Bayesian Hierarchical Models for Data Extrapolation and Analysis in Rare and Pediatric Disease Clinical Trials

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Cynthia Basu

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
Doctor of Philosophy

Advised by Bradley P. Carlin, Ph.D.

August, 2017

© Cynthia Basu 2017
ALL RIGHTS RESERVED

Acknowledgements

I must begin by expressing my sincere gratitude to my academic dad, Prof. Bradley P. Carlin, to whom I am forever in debt for his relentless patience, guidance, support, knowledge, and motivation. It was truly an honor to have gotten to know him not only as an astounding researcher, but as a wonderful human being and the most amazing advisor.

I would also like to thank my committee members Prof. Lin Zhang, Prof. Jim Hodges and Prof. Jim Cloyd not only for their support but also their invaluable inputs that widened my perspective on this work and helped me better understand my own topic.

I am also grateful to all my collaborators across the years, without whom none of this work would have been possible. I am thankful to my pharmacology collaborators who helped me comprehend and appreciate their subject. Their expertise and support is what enabled me to pursue problems in the area of pharmacokinetics/pharmacodynamics. I am especially grateful to my statistical collaborators at Amgen and the DIA Bayesian Statistics Working Group for all their hard work that went into developing these concepts. I am also thankful to my supervisors at the FDA for mentoring me for three months over the summer of 2015 and helping me better appreciate the practice of sensible regulation of clinical trials.

Finally, my heartfelt gratitude to my friends who always stood by me and made this journey more joyful. My fellow students in the Division of Biostatistics and at the university in general provided feedback, cooperation, and most importantly their friendship. Most of all I am thankful to my parents for believing in me and instilling in me their faith in education and knowledge.

Dedication

To my mother and father, Lucy and Sanjay Basu.

Abstract

A *rare disease* is defined by the Rare Diseases Act of 2002 as a disease that currently affects fewer than 200,000 patients in the USA. A pediatric population is one where the subjects are of age 18 or less. These two crucial yet underserved types of populations come with their own limitations in clinical trials. The paucity of potential trial enrollees and sensitivity of these patients, combined with a lack of sufficient natural history and experience, presents several economical, logistical and ethical challenges when designing clinical trials. An increasingly well accepted approach to address these challenges has been data extrapolation; that is, the leveraging of available data from adults or older age groups to draw conclusions for the pediatric population. Bayesian hierarchical modeling facilitates the combining (or "borrowing") of information across disparate sources, such as adult and pediatric data.

In this thesis we begin by developing, illustrating, and providing suggestions for Bayesian statistical methods that could be used to design improved clinical trials for pediatric and rare disease populations that efficiently use all available information. A variety of relevant Bayesian approaches are described, several of which are illustrated through two case studies: extrapolating adult binary efficacy data to expand the labeling for the drug Remicade® to include pediatric ulcerative colitis, and a simulated continuous longitudinal dataset patterned after an evaluation of the drug cinacalcet in treating pediatric secondary hyperparathyroidism (HPT).

The thesis then turns to methods useful in the study of X-linked adrenoleukodystrophy (X-ALD), a rare neurodegenerative disease for which Lorenzos Oil (LO) is one of the few available treatments. We offer a hierarchical Bayesian statistical approach to understanding the pharmacokinetics (PK) and pharmacodynamics (PD) of LO, linking its dose to the plasma erucic acid concentrations by PK modeling, and then linking this concentration to a biomarker (C26, a very long-chain fatty acid) by PD modeling. Next, we design a Bayesian Phase IIa study to estimate precisely what improvements in the biomarker can arise from various LO doses while simultaneously modeling a binary toxicity endpoint. Our Bayesian adaptive algorithm emerges as reasonably robust and efficient while still retaining good classical (frequentist) operating characteristics. Future work in this area looks toward using the results of this trial to design a Phase III study linking LO dose to actual improvements in health status, as measured by the appearance of brain lesions observed via magnetic resonance imaging.

Finally, the thesis shows how to utilize the rich PK/PD data to inform the borrowing of information from adults during pediatric drug development. Here we illustrate our approaches in the context of evaluating safety and efficacy of cinacalcet for treating HPT in pediatric and

adult patients. We use population PK/PD modeling of the cinacalcet data to quantitatively assess the similarity between adults and children, and in turn use this information in various hierarchical Bayesian rules for borrowing from adults, statistical properties of which can then be evaluated. In particular, we simulate the bias and mean square error performance of our approaches in settings where borrowing is and is not warranted to inform guidelines for the future use of our methods.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Data Extrapolation in Rare and Pediatric Diseases

A *rare disease* is defined by the Orphan Drug Act of 1983 as a disease that currently affects fewer than 200,000 patients in the USA. Disease rarity not only reduces the possible candidates and cases available for study and research, but also limits patient natural history and experience available. A *pediatric* disease is one that affects individuals aged 18 or less as define by section 529(a)(3) to the Federal Food, Drug, and Cosmetic Act (the FD&C Act). Studies involving drugs targeted at pediatric or rare diseases require special attention, as they present several economical, logistical and ethical challenges[26,15,7].

In the United States, legislative actions have been taken to promote research in these fields, such as the Best Pharmaceuticals for Children Act[6], Pediatric Research Equity Act[44], and the Orphan Drug Act[43]. As the populations involved are often more sensitive as well as fewer in number, we wish to minimize the risk involved while maximizing the little information we have. It would be very helpful if we could borrow information from all available data sources, both to help substantiate our results from the current study and to attempt to reduce the needed sample size. Borrowing historical data information has been well explored in regulatory science for medical devices, with guidance documents available from FDA[11,12]. However, borrowing information in the case of drugs is a not as straightforward, as the pharmacokinetics (PK) and the pharmacodynamics (PD) of the drug may be very different in, say, adults and children, and hence the response from children may be different from that of the adults[51]. Also, drugs are often approved for an adult population and prescribed off-label to children with only minimal understanding of the vagaries of pediatric response. Later, the sponsors often wish to conduct new studies to test the drug specifically in the pediatric population. In such studies it is unethical to subject more children to experimentation than absolutely required, motivating use of the

adult data from other clinical trials or any other historical dataset to strengthen the analysis and reduce the pediatric sample size. Many extrapolation techniques[10] have been used to streamline pediatric drug development and help speed approvals of drugs for pediatric labeling.

Bayesian statistical methods easily facilitate adaptive dose-finding and other early-phase studies, permit formal borrowing of strength from other information sources including expert opinion and previous data, and yield probabilistic inference regarding model quantities of interest. In this thesis we will look at Bayesian methods in several phases of clinical trials specifically related to rare and pediatric diseases, with emphasis on data extrapolation techniques to facilitate more efficient and ethical clinical trials.

We begin Chapter 2 with a look at hierarchical Bayesian models to evaluate and analyze Phase II-III rare and pediatric data. This is a particularly crucial area for borrowing of information from historical or adult datasets, since there is a need for Type I error control in these later, often regulated phases. We will look at various methods to borrow information, including power priors and commensurate priors. We will test methods using binary data from clinical trials pertaining to the drug Remicade (infliximab), which was under review by FDA as a treatment for pediatric unlcerative colitis. Next we formulate piecewise linear models for continuous longitudinal data and show how data extrapolation can be carried out with such models using a Bayes-Markov Chain Monte Carlo (MCMC) approach. We consider power and commensurate prior methods, as well as more standard methods as summarized by Thompson and Pennello[53]. We illustrate the methods to leverage data using a simulated dataset patterned after a real dataset we describe below.

Next, in Chapter 3 we expand our attention to the pharmacokinetics and pharmacodynamics (PK/PD) stage, and expand our hierarchical Bayesian modeling to drug PK/PD. Here we are motivated by a data set from a study on the effect of Lorenzo's Oil (LO) for patients with X-linked adrenoleukodystrophy (X-ALD), a rare disease of primarily young boys. Using information from our PK/PD modelling, we design a Phase IIa clinical trial for LO efficacy and toxicity that fine-tunes the dose administration per patient's personal response. In our design, we compare the use of standard Emax and approximating linear PD models, the latter of which can offer statistical computational benefits, especially when simulating trial operating characteristics.

Chapter 4 then introduces a novel method to utilize PK/PD data to inform the amount of borrowing of information from previous adult data during the later clinical phases of pediatric drug development. The basic idea here is that if the PK/PD properties of the drug are similar in adults and children, this helps justify borrowing more liberally from the adult data in later phases of the pediatric development program. Here again we tackle the case of longitudinal continuous data, and develop models for both efficacy alone, and efficacy and safety jointly. We

then use these models to analyze a dataset on pediatric secondary hyperparathyroidism (HPT), where our power prior is calibrated using PK/PD data from related adult and child studies. We then perform a simulation study to help shed light on the proper choice of calibration of our power prior. We close this chapter with preliminary results for a piecewise linear version of our longitudinal model, and some discussion of alternate methods for measuring adult-child similarity in the PK/PD data.

Finally, Chapter 5 concludes and summarizes the work. In particular, Section 5.1 focuses on direct extensions of the methods in Chapters 2-4, while Section 5.2 proposes a modified version of Zellner's g prior and describes its use in data borrowing from historical and adult datasets. We also analyse simulated Gaussian data based on the Remicade example using hierarchical Bayesian modeling, and again incorporating tools from borrowing methods such as power priors and commensurate priors. We also propose to develop better methods for selecting g in this setting.

1.2 Datasets Used in the Thesis

1.2.1 Remicade for Pediatric Ulcerative Colitis

In Section 2.2 we consider a beta-binomial model for binary response data illustrated with a recent dataset on the drug Remicade® (infliximab), a drug approved by the FDA to treat Crohn's disease, ulcerative colitis (UC), rheumatoid arthritis, and other conditions. We have data from two adult UC clinical trials, called ACT 1 and ACT 2, as well as one pediatric UC clinical trial called T72. Both ACT 1 and ACT 2 had patients who were administered infliximab at a dose 5mg/kg. The pediatric trial T72 had only 60 pediatric patients who were administered infliximab at 5mg/kg. This uneven sample size is often prevalent in such clinical trials, where the drug has been tested in adults previously and should be treated with caution.

Our endpoint of interest here is based on the observed Mayo score which is a commonly used activity index in placebo-controlled clinical trials for UC. The score is further elaborated in Chapter 2. The binary nature of the dataset as well as the uneven sample size across the trials present an interesting case study for the use of Bayesian data extrapolation methods in clinical trials. The dataset is further elaborated in the chapter along with its analysis.

1.2.2 Lorenzo's Oil for X-ALD

In Chapter 3, we analyze a PK/PD data from young boys suffering from X-ALD. We are interested in the effect of Lorenzo's Oil (LO) administration on the change in C26:0 plasma concentration in these boys, while adjusting for various covariates specific to individuals and

observations, such as the dose of LO administered, the weight of the patient, etc. LO is a 4:1 mixture of oleic and erucic acid, and administration of LO is known to increase levels of erucic acid in the plasma.

The analyzed dataset contains data from 116 subjects screened for this study at the John Hopkins Research Hospital from 2000 to 2014 under an expanded access trial (ClinicalTrials.gov, NCT02233257). Diagnosis of X-ALD was confirmed by plasma SVLFCA assay. These asymptomatic X-ALD children were followed until they developed any brain MRI abnormality. The patients returned to the clinic every few weeks for their blood samples to be collected, which resulted in one observation per visit. The lack of detailed observation per dose cycle of the patients presents a limitation to the usual PK/PD analysis. For the PK/PD analysis we are interested in the patients' erucic acid concentration, C26:0 level obtained from the blood samples and also each patient's weight in kg. Since the patients display elevated levels of C26:0, the average improvement in C26 level for each patient is the primary quantity of interest in our PD analysis as well as the Phase IIa trial. We look at further details and analysis in Chapter 3.

1.2.3 Cinacalcet for Pediatric Secondary HPT

We have two sets of data from clinical trials for cinacalcet, namely a dataset for a PK/PD analysis, and a Phase III clinical trial dataset.

Data for PK/PD Analysis

The dataset for PK/PD analysis consists of data from 7 clinical studies, 3 of which are adult studies (Amgen studies 20000172, 20000187 and 980126), and the remaining 4 pediatric studies (Amgen studies 20070208, 20110100, 20030227 and 20090005). The adult studies include a Phase III randomized controlled study with 403 patients (20000172), a Phase II study with 60 patients (980126), and a Phase I study with 22 patients (20000187). The pediatric studies include 2 studies with older pediatric patients aged 6-18 years (20070208, a Phase 3 randomized controlled study with 43 patients, and 20030227, a Phase I study with 12 patients), and 2 studies with younger pediatric patients aged 28 days to less than 6 years (20110100, a Phase 2 study with 11 subjects, and 20090005, a Phase 1 study with 12 patients). Due to the sequential nature of the PK/PD analysis explained later in the section, we have the posterior estimates of the PK analysis necessary for the PD analysis. We also have extensive patient specific observations necessary for the PD analysis, including the iPTH and cCa observations which constitute the crucial feedback loop in the model.

Data for Phase III Efficacy and Toxicity Analysis

The clinical efficacy and toxicity response data are further analyzed in Section 4.3. These data came from three adult Phase III clinical studies (20000172, 20000183 and 20000188) and one pediatric Phase III study (20070208). The three adult studies have similar study designs and were randomized, placebo-controlled clinical trials with subjects aged at least 18 years with secondary HPT receiving dialysis. Total sample size in these adult studies is N=1136. The pediatric data were collected from the 30-week, double-blind, placebo-controlled phase of pediatric study 20070208. The study enrolled N=43 subjects aged 6 to 18 years with secondary HPT receiving dialysis. During the course of the studies, a cinacalcet dose was given daily, and labs including iPTH and cCa were measured every two weeks. The percent change in the level of iPTH indicates the efficacy of the drug, whereas the drop in the level cCa below a threshold value is an indicator of toxicity in the patient.

Chapter 2

Hierarchical Bayesian Models for Clinical Data Extrapolation

2.1 Background

As mentioned in Chapter 1, statistical borrowing of strength for auxiliary data can offer a way forward in rare and pediatric drug and device approvals. In this chapter, we lay out such models for both continuous and binary data, and illustrate their application to two rare and pediatric disease settings.

2.1.1 Classical Statistical Approaches to Data Extrapolation

The traditional approach to analyzing a trial for pediatric diseases, where the drug has already been approved for adults, is to carry out a study on children and analyze it without any information borrowed from adult data. Depending on the type of dataset at hand, as well as the background clinical information, a wide variety of models can be fit to explain the dose-response relationship. These could be simple linear models, random or mixed effects survival models, piecewise linear regression models, or logit models for binary response datasets. Classical approaches include fitting simple frequentist random or mixed effects models to just the pediatric trial data, which is now routinely done using software packages such as `R` and `SAS`. An example for such a model is given in (2.1), where we regress the mean response μ_{ij} from the i^{th} child's j^{th} observation on the time from baseline, denoted here by $Time_{ij}$, the dose of the drug, denoted by $Dose_{ij}$, and introduce a subject-specific random intercept γ_i:

$$\mu_{ij} = \gamma_i + \beta_1 Time_{ij} + \beta_2 Dose_{ij} \ . \tag{2.1}$$

For a continuous observation Y_{ij}, a sensible likelihood is

$$Y_{ij} \overset{ind}{\sim} N(\mu_{ij}, \sigma_Y^2),\ i = 1, ..., n_c,\ j = 1, ...m_i, \tag{2.2}$$

for some $\sigma_Y^2 > 0$ where n_C is the number of children. However, this approach ignores the information in the adult dataset.

An alternate but still naive approach would be to pool all adult and child observations into one big dataset and then fit a model. This uncritical pooling of the pediatric and adult datasets may be inappropriate in many settings. A slight improvement to this procedure is to fit the regression once to the adult data and again to the pediatric data, and then take an appropriate weighted average of the pediatric and adult results. These weights could be based on the sample sizes of the two datasets, or perhaps expert opinion about the drug and the disease. However, the weight selection is crucial but difficult to justify in a traditional statistical framework.

A key aspect of such problems is the basis upon which extrapolation is deemed appropriate. Several pieces in the literature have discussed this issue[10], where a systematic review of approaches for matching adult systemic exposures is often used as the basis for dose selection in pediatric trials submitted to the FDA. The literature refers to two categories of extrapolation, *full* and *partial* extrapolation. Full extrapolation is when adult data are used directly to establish pediatric safety or efficacy. These extrapolations rely on data supporting the assumptions that there are similar disease progressions, responses to intervention, and exposure-response relationships in the adult and pediatric populations. Pediatric development supported by pediatric pharmacokinetic and safety data or pediatric safety data alone can be considered adequate. By contrast, partial extrapolation is when adult data are statistically combined with pediatric data to make such determinations. Partial extrapolation of efficacy is used when there is uncertainty about at least one of the assumptions underlying complete extrapolation, as mentioned above. In such cases, pediatric development could be based on a PK-PD study to confirm response in the pediatric population, followed by a single, adequate, well-controlled pediatric trial to confirm the efficacy seen in adults.

2.1.2 Current Bayesian Approaches

The approaches mentioned so far work well when we have a fairly rich pediatric dataset and also understand the disease and drug mechanisms reasonably well. However when we are working with pediatric or rare diseases, we wish to use every bit of information available to us. In such cases, Bayesian methods can help strengthen our analysis through their ability to combine multiple information sources. Current Bayesian approaches include fitting hierarchical models[5,8,30] to both the adult and pediatric data together, where the parameters connecting the two datasets

are related at some deeper level of the hierarchy. As an illustration, let us assume we have three datasets: two adult datasets D_1 and D_2, and a pediatric dataset D_0. Let the numbers of patients in each of these datasets be n_1, n_2, and n_0, respectively, and let $i = 1, ...n_1, ...(n_1 + n_2), ..., n$ denote the patient index, where $n = n_1 + n_2 + n_0$ denotes the total number of patients. Then assuming a simple linear fixed effects model like the one in (2.1), having likelihood (2.2), we can construct the following hierarchical model:

$$\begin{aligned}
\mu_{ij} &= \gamma_{a,i} + \beta_{a,1} Time_{ij} + \beta_{a,2} Dose_{ij} \text{ for } i = 1, ..., (n_1 + n_2) \\
\mu_{ij} &= \gamma_{c,i} + \beta_{c,1} Time_{ij} + \beta_{c,2} Dose_{ij} \text{ for } i = (n_1 + n_2) + 1, ..., n, \\
\gamma_{a,i} &\sim N(\gamma_{0a}, \sigma_a^2),\ \gamma_{c,i} \sim N(\gamma_{0c}, \sigma_c^2),\ \gamma_{0a}, \gamma_{0c} \overset{ind}{\sim} N(\gamma_0, \sigma_\gamma^2) \\
\text{and } \beta_{a,k}, \beta_{c,k} &\overset{ind}{\sim} N(\beta_k, \sigma_{\beta_k}^2) \text{ for } k = 1, 2.
\end{aligned} \tag{2.3}$$

We may proceed to assign hyperpriors to $\gamma_0, \beta_1, \beta_2$, and the variance parameters based on prior information from other data, such as crude auxiliary estimates of these numbers if possible, or using vague priors that let the data direct the values of these parameters. Such hierarchical models are widely used and often yield sensible results. Such models can however be constructed to incorporate much more prior information than shown above, leading the hierarchy to become more complicated and possibly rely too much on the adult data. Viele[58] and Hobbs *et al.*[18] also stress that such "static borrowing" (where the amount of adult data incorporated is fixed ahead of time) can lead to biased estimates when the two data sources do not agree. Also, when we rely on vague hyperpriors, the model may be computationally improper and our MCMC algorithm may fail to converge, thus rendering the Bayesian approach futile. Care should thus be taken to incorporate all prior information and expert opinion regarding the problem at hand, as well as to be reasonably parsimonious with parameters. Note also that all borrowing is implicit through the exchangeability of the $\beta_{a,k}$ and $\beta_{c,k}$. Thus this method is simple, but doesn't allow us to explicitly control the amount of borrowing between the datasets.

An alternative to fitting an exchangeable model for the datasets together is to take a two-step approach, where we first fit a hierarchical model to the adult dataset, and then use its posterior estimates as the priors in a second statistical model for the pediatric dataset. This method is illustrated in a real-life setting in Subsection 2.3.1. We might also downweight the adult information used by introducing various scaling parameters in the priors for the pediatric model.

In the longitudinal continuous data setting, suppose we first fit the adult model to our adult data and obtain posterior means and variances for γ_{0a}, $\beta_{a,1}$ and $\beta_{a,2}$, which we denote informally by $\hat{\gamma}_{0a}$, $\hat{\sigma}^2(\hat{\gamma}_{0a})$, $\hat{\beta}_{a,1}$, $\hat{\sigma}^2(\hat{\beta}_{a,1})$, $\hat{\beta}_{a,2}$, and $\hat{\sigma}^2(\hat{\beta}_{a,2})$. Then in Step 2, we assume $\gamma_{0c} \sim N(\hat{\gamma}_{0a}, \alpha_0 \hat{\sigma}^2(\hat{\gamma}_{0a}))$, $\beta_{c,1} \sim N(\hat{\beta}_{a,1}, \alpha_1 \hat{\sigma}^2(\hat{\beta}_{a,1}))$, and similarly for $\beta_{c,2}$ using α_2 to scale

the prior variance. The three hyperprior variance scaling parameters (α_0, α_1 and α_2) help us control the amount of borrowing from the adult data. In other words, we center our priors for these parameters at the posterior estimates we have obtained for the adults, but downweight this information by choosing α's bigger than 1. When these values equal 1, we assume full borrowing from the adult information.

A slightly more sophisticated and flexible method to facilitate data extrapolation while retaining control on the amount of borrowing is through the use of the *power prior*[25]. This approach downweights the supplemental (adult) likelihood by raising its likelihood in the posterior calculation to a power that is between 0 and 1. Assuming we wish to regress our response variable on some independent variables and the parameter of interest is θ, the power prior based on two adult datasets is

$$\pi(\theta, \boldsymbol{\lambda}|D_1, D_2) \propto \left[\prod_{k=1}^{2} L(\theta|D_k)^{\lambda_k}\right] \pi(\theta) \ , \tag{2.4}$$

where $\boldsymbol{\lambda} = (\lambda_1, \lambda_2)$ and the initial prior $\pi(\theta)$ is often vague. Note that λ_k controls how much information will be borrowed from auxiliary (adult) dataset k to supplement the (fully-utilized) child data; e.g., $\lambda_k = 1$ means full borrowing from source k, while $\lambda_k = 0$ implies no borrowing from this source. Such control is important in cases where there is heterogeneity between the supplemental and primary data, or when equal weighting of primary and all supplemental data sources is inappropriate. In fact, for fixed power priors, there is a one-to-one relationship between the power parameter and the effective sample size in the prior. The relationship is particularly straightforward in the normal likelihood setting; see for example Morita *et al.*[33,34] and Penello and Thompson[45].

Finally, the *commensurate* prior approach[18,19,20] is an even more fully adaptive method to account for the commensurability between the adult and pediatric datasets. It essentially specifies a hierarchical model with posterior

$$p(\theta_c, \theta_a, \eta|D_0, D_1, D_2) \propto L(\theta_c|D_0)L(\theta_a|D_1, D_2)\pi(\theta_c|\theta_a, \eta)\pi(\theta_a)\pi(\eta) \ . \tag{2.5}$$

Hence the so-called commensurate prior for the child parameter vector θ_c, $\pi(\theta_c|\theta_a, \eta)$, is usually centered around the corresponding parameter for the adult data (e.g., $N(\theta_c|\theta_a, \eta^{-1})$). The amount of borrowing can be modified by tuning the precision η of the prior around θ_a. A larger variance would imply we have less faith in the similarity of the pediatric and adult data, and therefore allow the pediatric estimate to be farther from that of the adult. Hobbs *et al.*[20] recommend a "spike and slab" hyperprior for η, which helps crystallize the choice between borrowing and not borrowing. Both this method and the power prior method are also demonstrated in our Subsection 2.3.2 case study.

Endpoint	ACT 1 (adult) Infliximab 5mg/kg $n = 121$	ACT 2 (adult) Infliximab 5mg/kg $n = 121$	T72 (pediatric) Infliximab 5mg/kg $n = 60$
Clinical response	84 (69.4%)	78 (64.5%)	44 (73.3%)
Clinical remission	47 (38.8%)	41 (33.9%)	24 (40.0%)
Mucosal healing	75 (62.0%)	73 (60.3%)	41 (68.3%)

Table 2.1: Study-level endpoint data, Remicade UC studies in adults (ACT 1 and ACT 2) and pediatrics (T72)

All Bayesian models described above can be fit using standard Bayesian software such as `OpenBUGS`[31], `Stan (mc-stan.org)`, `Proc MCMC` in `SAS`, or in `R` using packages available in `CRAN (cran.r-project.org)` or those that call `BUGS` or its variants from `R`, such as `rjags`.

2.2 Binary Response Model

As discussed in Section 1.2.1, we look at a beta-binomial model for binary response data on the drug Remicade® (infliximab). Its maker sought a meeting with a gastrointestinal (GI) FDA advisory panel for the purpose of expanding the drug's labeling to include *pediatric* ulcerative colitis. As is common in such settings, extrapolation from adult data was not permitted for dosing or safety assessment in children, but the panel did allow the sponsors to argue for extrapolation of efficacy using two existing adult studies. Ultimately the panel did decide in favor of full extrapolation, but no quantitative modeling was involved in this decision, only clinical judgment.

The summary statistics of the data seen by the advisory panel have appeared in the literature[23,24]. Recapitulating from Section 1.2.1, our data comprises of $K = 2$ UC trials in adults (called ACT 1 and ACT 2) and one UC trial in pediatrics (called T72). The efficacy endpoints in these trials are based on the Mayo score derived from the subscores of its 4 components: stool frequency, rectal bleeding, endoscopic findings, and physician's global assessment. The Mayo score has a minimum value of 0 and a maximum value of 12. The primary efficacy endpoint of clinical response at week 8 is defined as a decrease in the Mayo score by at least 30% and 3 points, with a decrease in the rectal bleeding subscore of at least 1 point or a rectal bleeding subscore of 0 or 1. When this definition is met, the clinical response is 1; otherwise, it is 0. The secondary endpoints are presence or absence of clinical remission and mucosal healing at week 8. Table 2.1 gives the summary statistics of the endpoints. In the remainder of this section, we apply Bayesian binary methods to these data, obtaining quantitative summaries that might have helped the panel make a better-informed decision.

Two Step Prior		Commensurate Prior		Power Prior	
r (ESS)	$E(\theta_3\|D, D_0)$ (CI)	κ_α (ESS)	$E(\theta_3\|D, D_0)$ (CI)	α_0(ESS)	$E(\theta_3\|D, D_0)$ (CI)
0.01 (62.4)	0.729 (0.616,0.832)	1 (62.4)	0.730 (0.622,0.837)	0 (62)	0.725 (0.615, 0.835)
0.25 (121)	0.687 (0.607,0.761)	10 (84)	0.724 (0.620,0.828)	0.25 (123)	0.710 (0.631, 0.790)
0.5 (181)	0.667 (0.602,0.728)	50 (181)	0.705 (0.617,0.793)	0.5 (183)	0.705 (0.638, 0.770)
1 (302)	0.659 (0.606,0.711)	100 (302)	0.700 (0.622,0.778)	1 (304)	0.700 (0.648, 0.752)

Table 2.2: Results for the various Bayesian models fit to the study-level Remicade data.

2.2.1 Two-step Approach

Let Y_{jk} denote the binary outcome on the primary endpoint for patient j in study k, and $Y_k = \sum_{j=1}^{J} Y_{jk}$ denote the summary statistic for the primary endpoint total in study k, $k = 1, 2, 3$. Thus, the adult data is $\boldsymbol{D}_0 = (Y_1, Y_2)$ and the pediatric data is $D = Y_3$. We assume that $Y_k \sim \text{Binomial}(n_k, \theta_0)$, for $k = 1, 2$ and use a conjugate $\text{Beta}(\kappa_a\mu_a, \kappa_a(1-\mu_a))$ prior on θ_0. We complete the model specification by assigning hyperpriors $\kappa_a \sim \text{Uniform}(2, 122)$ and $\mu_a \sim \text{Beta}(1, 1)$, where the upper bound for κ_a was chosen to be comparable to the sample size of either of our adult datasets. The posterior distribution for a given κ_a and μ_a is $\text{Beta}(\kappa_a\mu_a + \sum_{k=1}^{2} Y_k, \kappa_a(1-\mu_a) + \sum_{k=1}^{2}(n_k - Y_k))$. Note that the support of (κ_a, μ_a) is $(2, 122) \times (0, 1)$.

In step 1, given only the adult data D_0, denote the posterior means of κ_a and μ_a by $\hat{\kappa}_a$ and $\hat{\mu}_a$, and use these in the prior for the pediatric dataset. Specifically, we assume $\theta_3 \sim \text{Beta}(r\hat{\kappa}_a\hat{\mu}_a, r\hat{\kappa}_a(1-\hat{\mu}_a))$, where we use $r \in (0, 1)$ to scale down the adult data effective sample size (ESS), $\hat{\kappa}_a$, as the pediatric population (n_3=60) is much smaller than the combined adult populations ($n_1 + n_2 = 242$). Note that r can be either assumed known or have a $\text{Beta}(1, 1)$ prior; in the latter case is determined from the data. Here, we assumed that r is known. In step 2 of this two step approach, we update the $Binomial(n_3, \theta_3)$ likelihood for Y_3 to obtain a $\text{Beta}(r\hat{\kappa}_\alpha\hat{\mu}_\alpha + Y_3, r\hat{\kappa}_a(1-\hat{\mu}_a) + n_3 - Y_3)$ posterior for θ_3. Note that the empirical pediatric success rate is slightly higher than that of the adults, and hence when we borrow more strength from the adult data, the θ_3 estimates should decrease, and the corresponding 95% credible intervals should get narrower.

The results for various choices of r (corresponding to ESS values ranging from 62.4 up to the full combined sample size of 302) are given in the first two columns of Table 2.2. Notice that as we increase the r, the θ_3 point estimate decreases (from 0.729 down to 0.659) and the corresponding 95% credible interval widths also decrease (from 0.216 down to 0.105), both as expected. In this model, the choice of r is somewhat subjective, but values greater than 0.5 result in $60/181 = 33\%$ or less of θ_3 posterior's strength coming from the pediatric data, which may be insufficient for regulatory approval.

2.2.2 Combined Approaches

In this section we first fit a commensurate prior model. In particular, we choose a $\mathrm{Beta}(1,1)$ distribution as our initial prior on θ_0 and take $\theta_3|\theta_0 \sim \mathrm{Beta}(\kappa\theta_0, \kappa(1-\theta_0))$ as our commensurate prior, with $\kappa \sim \mathrm{Gamma}(\kappa_\alpha, 1)$ and κ_α assigned a fixed value. The commensurate prior approach then specifies a hierarchical model with posterior

$$p(\theta_3, \theta_0, \eta|D, D_0) \propto L(\theta_3|D)L(\theta_0|\boldsymbol{D}_0)\pi(\theta_3|\theta_0, \eta)\pi(\theta_0)\pi(\eta)\ . \tag{2.6}$$

Thus the joint posterior in this case arises as

$$\pi(\theta_3, \theta_0, \kappa|D, D_0) \propto \\ \theta_3^{Y_3}(1-\theta_3)^{n_3-Y_3}\theta_0^{Y_1+Y_2}(1-\theta_0)^{n_1+n_2-Y_1-Y_2}\theta_3^{\kappa\theta_0-1}(1-\theta_3)^{\kappa(1-\theta_0)-1}\kappa^{\kappa_\alpha-1}e^{-\kappa}. \tag{2.7}$$

While this does not lead to a closed form for the marginal posterior $p(\theta_3|D, D_0)$, sampling from the distribution is routine via the `BUGS` language. The results for varying values of κ_α are given in Table 2.2; the ESS values shown for this method are computed as functions of our posterior estimates. As in the case with the two-step approach, increases in κ_α are again associated with clear decreases in the θ_3 point estimates and interval widths, though the shrinkage back to the adult values is less dramatic here.

Finally, we also apply the power prior method to our data. Here we may begin by assuming that $\theta_0 = \theta_3 = \theta$, instead of putting a homogeneity constraint on the θ_ks. We then obtain the posterior $p(\theta|D, \boldsymbol{D}_0)$ as proportional to $L(\theta|D)L(\theta|\boldsymbol{D}_0)\pi(\theta)$. The power prior approach[25] downweights the supplemental (adult) likelihood by raising it to a power α_0 that is between 0 and 1. The power prior then arises as

$$\pi(\theta, \boldsymbol{\alpha}_0|\boldsymbol{D}_0) \propto \left[\prod_{k=1}^{2} f(Y_k|\theta)^{\alpha_{0,k}}\right]\pi(\theta)\pi(\boldsymbol{\alpha}_0)\ . \tag{2.8}$$

Note that $\alpha_{0,k}$ controls how much information will be borrowed from the auxiliary (adult) data to supplement the fully-utilized child data; e.g., $\alpha_{0,k} = 1$ means full borrowing from source k, while $\alpha_{0,k} = 0$ implies no borrowing. Such control is important in cases where there is heterogeneity between the supplemental and primary data, or when equal weighting of primary and supplemental samples is inappropriate. In fact, if the power priors are fixed, there is a one-to-one relationship between the power parameter of the power prior and the variance of the prior; the relationship is particularly straightforward in the normal likelihood setting.

We fix the powers $\alpha_{0,k} = \alpha_0$ for $k = 1, 2$, thus specifying a fixed and equal amount of borrowing from both adult studies. Notationally, we have $Y_k \sim \mathrm{Binomial}(n_k, \theta_3)$ and $\theta_3 \sim \mathrm{Beta}(\kappa\mu, \kappa(1-\mu))$, where we choose minimally informative values $\kappa = 2$ and $\mu = 0.5$. Then

$\theta_3|D, D_0 \sim \text{Beta}(\tilde{\kappa}\tilde{\mu}, \tilde{\kappa}(1-\tilde{\mu}))$, where $\tilde{\kappa} = \kappa + n_3 + \sum_{k=1}^{2} n_k\alpha_0$ and $\tilde{\mu} = \tilde{\kappa}^{-1}(\kappa\mu + Y_{K+1} + \sum_{k=1}^{K} Y_k\alpha_{0,k})$. The results for four representative values of α_0 (and corresponding ESS values) are given in the last column of Table 2.2. The effect of increasing ESS is again apparent, with the now-familiar trends in the posterior means and interval widths being somewhat intermediary to those arising from the previous two methods.

2.3 Continuous Longitudinal Response Models

Throughout this section our interest is driven by the cinacalcet data described in Section 1.2.3. Mimicking this dataset, we simulate datasets from both adult and pediatric clinical studies of Cinacalcet in the context of a linear mixed effects Bayesian hierarchical model to study the drug's effect on iPTH. Cinacalcet has been shown to lower the parathyroid hormone (PTH) released by the parathyroid glands, which in turn reduces the level of calcium and phosphorous released from the bones. The iPTH (intact PTH) test level is of key interest, and is routinely monitored for people with chronic kidney disease; lowering it by some clinically significant percentage is a goal of many efficacy trials in this area. The number of patients in the pediatric dataset is assumed to be just $n_c = 40$, whereas we suppose there are $n_a = 800$ patients in the adult study, a level of imbalance not uncommon in practice. Let $X_{i,j}$ be the percent change in the iPTH level of patient i ($i = 1, ..., n_c, ..., (n_c + n_a)$) in the week of the patient's j^{th} observation, $j = 1, ..., m_i$ (where m_i varies from 3 to 25). That is:

$$X_{i,j} = \frac{\text{iPTH}_{i,j} - \text{baseline iPTH}_i}{\text{baseline iPTH}_i} \times 100\ . \tag{2.9}$$

This percentage change will be our outcome variable in the linear model. Let $t_{i,j}$ denote the week after baseline for the j^{th} observation on the i^{th} patient. Since here we assume we don't have precise dosing information for every patient, our model (2.1) for the children becomes

$$\begin{aligned} X^c_{i,j} &\sim Normal(\mu^c_{i,j}, 1/\tau^c_e),\ i = 1, ..., n_c \\ \text{where } \mu^c_{i,j} &= \mu^c_{1i}t^c_{i,j} + I(drug^c_i = 1)(\mu^c_d t^c_{i,j}). \end{aligned} \tag{2.10}$$

Here μ^c_{1i} are the subject-level random effects, assumed to independently follow a $N(\eta^c_0, \tau^c_\eta)$ specification, and μ^c_d is the fixed effect of the drug on each child's slope. Note we do not include intercepts in model (2.11) since X_{ij} is defined to be 0 at baseline ($t_{ij} = 0$). Similarly for the adults, we assume

$$\begin{aligned} X^a_{i,j} &\sim Normal(\mu^a_{i,j}, 1/\tau^a_e),\ i = n_c + 1, ..., n_c + n_a \\ \text{where } \mu^a_{i,j} &= \mu^a_{1i}t^a_{i,j} + I(drug^a_i = 1)(\mu^a_d t^a_{i,j}). \end{aligned} \tag{2.11}$$

Now the μ_{1i}^a are subject-level random effects, assumed to independently follow a $N(\eta_0^a, \tau_\eta^a)$ specification, whereas μ_d^a is the fixed effect of the drug on the slope of the fitted iPTH percent change variable. Regarding hyperpriors, both η_0^c and η_0^a are assigned flat hyperpriors, μ_d^c and μ_d^a are assigned vague normal priors, and we place vague $G(0.1, 0.1)$ hyperpriors on τ_e^c and τ_e^a.

2.3.1 Two-Step Approach

We first fit a two-step model along the lines of those described in Section 2.1.2. Specifically, in Step 1 we fit the adult model to our adult data and obtain posterior means and variances for η_0^a, and μ_d^a, which we denote by $\hat{\eta}_0^a$, $\hat{\sigma}^2(\hat{\eta}_0^a)$, $\hat{\mu}_d^a$, and $\hat{\sigma}^2(\hat{\mu}_d^a)$. Then in Step 2, we use these posterior estimates to guide our pediatric analysis. Specifically, we assume $\eta_0^c \sim N(\hat{\eta}_0^a, \alpha_0 \hat{\sigma}^2(\hat{\eta}_0^a))$ and similarly $\mu_d^c \sim N(\hat{\mu}_d^a, \alpha_d \hat{\sigma}^2(\hat{\mu}_d^a))$. The next question is therefore what the values of α_0 and α_d should be. They can be assigned based on expert knowledge, such as how similar clinicians think the two populations are likely to be and thus how much borrowing can be justified. As mentioned above, such static borrowing is straightforward but clearly somewhat subjective. Alternatively, we can assign hyperpriors to the α's, such as a vague gamma or "spike and slab" distribution. However in some cases this may not be a good idea as it is often difficult to specify this hyperprior, and no information in our data exists to inform this decision.

A quantity sometimes used to guide this decision is the effective historical sample size (EHSS)[19], which generalizes our notion of ESS in the beta-binomial case of Section 2.2. Various definitions exist, but a straightforward one[53] for a parameter of interest ξ is

$$EHSS(\xi) = n_c \left[\frac{Var(\xi|X^c)}{Var(\xi|X^c, X^a)} - 1 \right], \tag{2.12}$$

which is the percent improvement in ξ's posterior precision (inverse variance) arising from using the chosen fraction λ of the adult data, expressed on the same scale as the child sample size. Note that other metrics (e.g., $\sqrt{Var^{-1}}$) might have been chosen in (4.1), and that the answer will vary with the choice of ξ.

In our case, we take the overall fitted slope in the drug group, $\eta_0^c + \mu_d^c$, as the primary parameter of interest ξ. Ideally the effective historical sample size should be no greater in magnitude than the pediatric sample size, since even though we are trying to borrow strength from the adult data, our analysis should be primarily driven by the pediatric data. However for our simulated data we often see EHSS values approaching or even exceeding the actual adult sample size of 800, since definition (4.1) is simple but may perform erratically in more complex hierarchical models, especially when implemented via MCMC.

Tables 2.3–2.5 contain the results of the methods applied to our simulated data. We look at the estimated posterior mean and standard deviation (sd), the 95% equal tail Bayesian credible

	η_0^c					$\eta_0^c+\mu_d^c$				
α	mean	sd	BCI lower	BCI upper	EHSS	mean	sd	BCI lower	BCI upper	EHSS
1	0.67	0.35	-0.04	1.35	1139.02	-3.12	0.53	-4.15	-2.08	703.1
7	1.13	0.86	-0.59	2.78	102.17	-2.60	1.24	-5.05	-0.19	96.94
15	1.54	1.14	-0.76	3.74	73.02	-2.24	1.59	-5.39	0.86	42.86
100	3.05	1.92	-0.76	6.72	0	-1.76	2.29	-6.31	2.69	0

Table 2.3: Posterior estimates for the model coefficients using a two-step approach for various values of α.

interval (BCI), and the calculated EHSS corresponding to the placebo effect, η_0^c, and the overall slope in the treatment group, $\eta_0^c+\mu_d^c$. The latter indicates whether the patients in the treatment group showed improvement over time.

In Table 2.3 we fit our two-step model using three different values of $\alpha = \alpha_0 = \alpha_d$ to show the effect of borrowing. The posterior estimates from Step 1 were $\hat{\eta}_0^a = 0.5659$ and $\hat{\mu}_d^a = -3.805$. Note that as α increases, the estimates for the children become more dissimilar to those of the adults. Note also that for the vaguest prior (α=100), we obtain the child data-only results (EHSS=0), which also have the largest estimated sd's. It appears for around $\alpha = 15$, $EHSS(\eta_0^c + \mu_d^c)$ is fairly close to the actual size of the pediatric dataset ($n_c = 40$). It should also be noted that based on the upper limit of the BCI for $\eta_0^c + \mu_d^c$, for example, the significance of our findings for the primary parameter change with α. In our case the change happens around $\alpha = 7$; smaller values lead to statistically significant findings, whereas larger values do not.

2.3.2 Combined Approaches

In this subsection we begin by fitting a commensurate prior model. We assume our model for the adults is as defined previously. We also model the pediatric data similar to before, but modify its prior to adaptively learn from the adult data based on their estimated similarity. Following (2.6), we assign the prior for η_0^c as $N(\eta_0^a, 1/\tau_c)$, and the prior for μ_d^c as $N(\mu_d^a, 1/\tau_{dc})$ to introduce the commensurability. We then proceed to assign spike and slab hyperpriors on τ_c and τ_{dc} as follows:

$$\tau_c \sim \begin{cases} Normal(200, 0.01) & \text{with probability } f; \\ Uniform(0.1, 5) & \text{with probability } 1-f, \end{cases}$$

$$\text{and } \tau_{dc} \sim \begin{cases} Normal(200, 0.01) & \text{with probability } f; \\ Uniform(0.1, 5) & \text{with probability } 1-f, \end{cases}$$

	η_0^c					$\eta_0^c+\mu_d^c$				
p	mean	sd	BCI lower	BCI upper	EHSS	mean	sd	BCI lower	BCI upper	EHSS
1	0.58	0.15	0.29	0.87	7827.4	-3.23	0.15	-3.53	-2.94	9464.6
0.9	0.75	0.77	0.29	3.57	244.54	-3.02	0.93	-3.61	0.15	211.08
0.65	1.33	1.50	0.19	5.45	35.36	-2.47	1.66	-5.13	1.66	38.43
0.5	1.77	1.78	0.14	6.11	13.73	-2.18	1.91	-5.73	1.98	19.35
0	3.28	2.06	-0.72	7.38	0	-1.72	2.32	-6.36	2.83	0

Table 2.4: Posterior estimates for the model coefficients using the commensurate prior for various values of p.

	η_0					$\eta_0+\mu_d$				
λ	mean	sd	BCI lower	BCI upper	EHSS	mean	sd	BCI lower	BCI upper	EHSS
0.1	0.78	0.16	0.46	1.09	8504.4	-3.01	0.16	-3.32	-2.71	10706.6
0.001	2.24	1.11	0.07	4.40	137.31	-2.57	1.12	-4.72	-0.33	169.99
0.00015	3.23	1.68	-0.09	6.52	37.22	-2.11	1.76	-5.52	1.48	44.43
0	4.11	2.34	-0.392	8.84	0	-1.46	2.56	-6.46	3.47	0

Table 2.5: Posterior estimates for the model coefficients using the power prior for various values of λ.

where we assume a single spike probability $f \sim Bernoulli(p)$ for both τ_c and τ_{dc}. We can now vary the amount of borrowing by varying the value of p. An increase in p would mean a higher chance of the precision taking a value close to 200, our "spike", and hence more borrowing from the adult data. By contrast, small values of p encourage small τ values in the "slab", which discourages adult borrowing. Relatively little information on τ_c and τ_{dc} exist in the data, so the spike and slab parameters must be tuned carefully.

Table 2.4 contains the results of this model for varying values of p. As can be seen, the results show the expected trends regarding borrowing between the datasets. For $p = 0$, we again get values consistent with no borrowing (EHSS=0). The $EHSS(\eta_0^c + \mu_d^c)$ indicates a p of 0.65 delivers an EHSS approximately the size of the pediatric dataset. However the EHSS does increase greatly for $p > 0.9$, becoming more than the available number of adult patients ($n_a = 800$). As noted in the case of the two-step approach, the significance of our findings changes with an increase in p, reemphasizing the caution with which the value of these parameters should be chosen. For $\eta_0^c + \mu_d^c$, the change seems to occur around p approximately 0.9; the BCI contains zero for smaller p, and does not contain zero for very large p.

Finally we fit a power prior model. Here we may begin by assuming that $\eta_0^a = \eta_0^c = \eta_0$ and $\mu_d^a = \mu_d^c = \mu_d$. We then obtain the posterior $p(\mu_d|D_a, D_c)$ as proportional to $L(\mu_d|D_c)L(\mu_d|D_a)^\lambda\pi(\mu_d)$.

For this method the EHSS was especially sensitive to the choice of λ. It appears that a very small λ of 0.00015 is needed to yield an EHSS close to the size of the pediatric dataset. The BCI for $\eta_0 + \mu_d$ again shows that the significance of our findings change with a change of λ. The upper boundary of the BCI for $\lambda = 0.001$ is close to zero, with this value roughly forming the break between Bayesian significance and insignificance.

If we fit a model as in the first step of the two-step approach to the pediatric data alone, the posterior estimate of η_0^c is 4.518 and $\eta_0^c + \mu_d^c$ is –2.02. As can be seen, for all 3 methods in Tables 2.3–2.5 as we move down the columns, the amount of borrowing from the adult data decreases and the obtained posterior estimates are now closer to those obtained by a hierarchical model on the pediatric data alone. Since p in the case of commensurate priors and λ in case of power priors vary from 0 to 1, it is easier to intuitively set their values in accordance with our prior knowledge. EHSS performance is somewhat unstable in Table 2.5, suggesting the commensurate approach may be preferable. If we wish to limit the EHSS to roughly the pediatric sample size ($n_c = 40$), we must conclude the drug does not lead to a significant improvement in iPTH percent change under any of our three methods. However if we can tolerate a higher EHSS (say, greater than $5n_c$), for example due to added clinical justification or relaxed restrictions on borrowing, we can arrive at significant results and conclude that the drug shows improvement in pediatric populations as well.

We also investigated individual mean levels for two individuals in our simulated dataset, numbers 35 and 9. The 35^{th} individual had iPTH observations larger than the other pediatric patients, whereas the 9^{th} patient was more or less randomly chosen and is not outlying. In both cases, EHSS could not be reliably estimated due to Monte Carlo error and the inherent smallness of these values. Thus EHSS does not seem to be a helpful tool in the case of individual random effects.

2.4 Outlook

As demonstrated by our Section 2.2 and Section 2.3 examples, several Bayesian modeling techniques exist to facilitate borrowing information between datasets for which we can control the degree of borrowing. Caution should be exercised when using these methods, especially when the datasets are very dissimilar and the adult or other historical dataset has the potential to sway the answers overmuch. That being said, given sufficient familiarity with these methods and corresponding software, they can be used in a wide variety of problems to strengthen analysis, especially in studies of pediatric and rare disease where we have to work with smaller sample sizes for ethical and logistical reasons. As noted above, with an increase in α, decrease in p or

decrease in λ, the statistical significance (i.e., whether the BCI includes zero), reemphasizing that the question of how to pick the degree of borrowing often remains hard to answer. The amount of borrowing can be based on several factors, such as expert opinion, or similarity between the adult and pediatric datasets. Our methods provide a way to quantify the trade-off between obtaining significant findings and fully justified adult data borrowing. However, more experience is needed with the spike and slab and other hyperpriors that control the degree of borrowing.

Chapter 3

Bayesian Models for PK/PD and Phase IIa Studies in Rare Diseases

3.1 Motivating Example in Rare Disease

X-linked adrenoleukodystrophy (X-ALD) is a rare (incidence of 1:17,000 [37]), progressive neurodegenerative disease which affects cerebral white matter, peripheral nerves, adrenal cortex and testis [35] that typically strikes boys aged 3 to 14. It is caused by mutation in the *ABCD1* gene that codes for a peroxisomal transporter [40,4]. Biochemically the disease is characterized by elevated plasma and tissue levels of the saturated very long-chain fatty acids (SVLCFAs) C24:0 and C26:0 [36].

As described in Section 1.2.2, we model the effect of Lorenzo's Oil (LO) administration on the change in C26:0 plasma concentration in boys suffering from X-ALD. We also adjust for various covariates specific to individuals and observations, namely the dose of LO administered, the weight of the patient, etc. As mentioned, administration of LO is known to increase levels of erucic acid in the plasma. Several papers have shown that erucic acid can normalize C26:0 plasma levels within 4-6 weeks[3,48]. However, little has been done to establish its clinical efficacy or indications for its use [1]. Although several open-label clinical trials (i.e., trials where both the researchers and participants knew which treatment was being administered) were unable to establish the clinical benefit of LO in patients with childhood cerebral adrenoleukodystrophy (CCALD) [2,56], other studies have indicated that LO *can* prevent CCALD when treatment is administered prior to cerebral involvement [39,38], and there are also indications of a benefit in the smaller adult subpopulation [27].

Hence our broad action plan for modeling can be summarized as:

$$LO \xrightarrow{PK} \text{Erucic Acid Concentration} \xrightarrow{PD} \text{change in C26 level} \longrightarrow \text{Brain MRI abnormalities}$$

In this diagram, the first link is captured by the PK modeling, while the second involves PD modeling, both of which we discuss below. Our Phase IIa design will permit us to hone in on a rationally-derived LO dose that accounts for both toxicity and efficacy (here, measured by the improvement in biomarker response).

We had access to data on 130 subjects screened for the study at the John Hopkins Research Hospital from 2000 to 2014 under an expanded access trial (ClinicalTrials.gov, NCT02233257). As noted, the diagnosis of X-ALD was confirmed by plasma SVLFCA assay. Fourteen subjects with incomplete data (i.e., missing either dose, or erucic acid, and/or C26:0 concentrations) were excluded from the study, resulting in $n = 116$ subjects for analysis who were followed until they developed any brain MRI abnormality. All participants received a daily dose of 2-3mg/kg of LO, a dose calculated to provide roughly 20% of caloric intake, along with maintaining some dietary restrictions.

Blood samples were collected at baseline as well as every month for the first 6 months, followed by collection at 3- to 6-month intervals after LO administration. The samples were processed and plasma was analyzed to obtain a fatty acid profile (FAP) for 70 fatty acids, including C26:0 and erucic acid. From the 116 subjects, 2384 paired C26:0 and erucic acid concentrations were available to develop the PD model. Hence we have multiple observations per patient, but only one observation per patient per visit. Usually in pharmacokinetic studies, we have data from samples collected at various time points after the drug was administered. This makes tasks such as deriving the half life of the drug, required for rational drug dosing strategies, straightforward. In our case we have only one observation per LO intake. With this less rich type of data, we must make more heroic assumptions in order to estimate the drug's half life.

3.2 Hierarchical Bayesian PK/PD Modelling

3.2.1 The Pharmacokinetic Model

Let Y denote the erucic acid concentration, and Z be the C26:0 level obtained from the FAP. We also have each patient's weight in kg, denoted here as wt, taken at each reading. For any drug, the average steady state concentration (css) can be written as the dosing rate divided by the *clearance*, which is the volume of the drug cleared totally from the blood (in our case, erucic acid) per unit time [9]. Our erucic acid observations (which we assume are at steady state) are actually comprised of two parts: the *endogenous* part (i.e., erucic acid produced by the body itself) and the *exogenous* part (erucic acid coming from LO administration). Hence for the jth observation on the ith patient, where $i = 1, ..., 116$ and $j = 1, 2, ...n_i$, we model clearance using

allometric scaling as

$$clearance\_erucic_{ij} = \theta_1 \left(\frac{weight_{ij}}{70} \right)^{\frac{3}{4}} . \tag{3.1}$$

Since the dose was given in terms of ml, we need to multiply it by the density of erucic acid, which is 0.860 g/ml (and therefore 860 mg/ml). Thus $Rate\_erucic_{ij} = \frac{860 \times Dose}{24}$. Next let $cssex_{ij} = Rate\_erucic_{ij}/clearance\_erucic_{ij}$, where $cssex$ denotes the exogenous css. We assume erucic acid concentration is at steady state. Let $cssav$ denote the average steady state concentration, $cssav_{ij} = cssex_{ij} + endo\_erucic_i$, where $endo\_erucic_i$ denotes the endogenous css for patient i. We model the observed css Y as this $cssav$ plus some random Gaussian error. Hence in statistical modeling terms, for Y_{ij}, the level of erucic acid observed in the FAP analysis of the jth sample on the ith patient, we have

$$\begin{gathered} Y_{ij} = cssav_{ij} + (error_{ij} e^{\eta_{3i}}) , \\ \text{where } error_{ij} = cssav_{ij} \epsilon_{1ij} \text{ and } \epsilon_{1ij} \sim N(0, \sigma_1^2) . \end{gathered} \tag{3.2}$$

Incorporating subject-level variability and utilizing equation (3.1), we obtain

$$clearance\_erucic_{ij} = \theta_1 \left(\frac{weight_{ij}}{70} \right)^{\frac{3}{4}} e^{\eta_{1i}} \text{ and } endo\_erucic_i = \theta_2 e^{\eta_{2i}} ,$$

where $\eta_{ki} \overset{indep}{\sim} N(0, \omega_k^2)$ for $k = 1, 2, 3$, the number of subject level errors we include in our model, $\theta_l \sim N(\mu_l, \tau_l^2)$ for $l = 1, 2$, and we put a vague Gamma(0.1,0.1) prior on $\frac{1}{\sigma_1^2}$.

3.2.2 The Pharmacodynamic Model

The $Emax$ model is a nonlinear model frequently used in pharmacodynamic dose-response analyses. In this set up, the basic model is

$$Response_{ij} = E_0 - \frac{Emax_i * cssex_{ij}}{EC50_i + cssex_{ij}} , \tag{3.3}$$

where $Emax$ denotes the maximum effect attributable to erucic acid concentration, $EC50$ denotes the erucic acid concentration that produces half of $Emax$, and E_0 is the basal effect, corresponding to the patient response when the erucic acid concentration is zero [32]. Since in our case the effect is decreasing with increasing concentration of the drug, we will use the minus sign[55,54] in equation (3.3). The observed effect is this response plus some error. Converting to a statistical model, for Z_{ij} (the level of C26 observed in the FAP analysis) we have

$$Z_{ij} = (C26endo_i - Effect_{ij})(1 + \epsilon_{2ij}) , \tag{3.4}$$

node	prior sd	mean	posterior sd	MC error	2.5%	median	97.5%
σ_1		0.002215	6.269E-5	8.32E-7	0.002093	0.002215	0.002341
σ_2		4.438	0.1509	0.003391	4.145	4.435	4.742
θ_1(clearance)	5	31.0	0.6158	0.01766	29.82	30.99	32.22
θ_2(endoerucic)	0.1	0.6083	0.1003	0.003036	0.4162	0.6139	0.8022
θ_3(Emax)	0.1	0.4573	0.02572	0.002346	0.3968	0.4591	0.5105
θ_4(EC50)	10	0.8931	0.6321	0.04938	-0.1921	0.8446	2.207
θ_5(C26endo)	0.1	0.8822	0.02605	0.002383	0.8256	0.8848	0.9313

Table 3.1: Table depicting parameter estimates obtained from the model described

where $\epsilon_{2ij} \overset{iid}{\sim} N(0, \sigma_2^2)$. To add subject-level variability to our model, we can extend (3.3) by defining $Emax_i = \theta_3 e^{\eta_{4i}}$, $EC50_i = \theta_4 e^{\eta_{5i}}$, and in (3.4) we can also assume $C26endo_i = \theta_5 e^{\eta_{6i}}$. As in the PK model, $\eta_{ki} \overset{indep}{\sim} N(0, \omega_k^2)$ for $k = 4, 5, 6$, the patient-level errors we include, and $\theta_l \sim N(\mu_l, \tau_l^2)$ for $l = 3, 4, 5$. Finally, we again add a vague prior, $\frac{1}{\sigma_2^2} \sim Gamma(0.1, 0.1)$

Using the data in hand, we seek to fit our Bayesian hierarchical model with some vague and some informative priors in an attempt to estimate all the unknown quantities. We also choose appropriate scales for the error components, both at the individual and observational levels. Our software platform is `WinBUGS`[31], which uses Markov chain Monte Carlo (MCMC) [8] methods to sample from the joint posterior distribution of all unknown model parameters.

3.2.3 Results for the LO PK/PD Data

After trying various modifications, due to weak identifiability the model converged best with no individual level errors, i.e., no η_{ki} terms. Our MCMC approach requires us to assign some starting values for the θ_{li}, $l = 1, ..5$, and also the σ_m^2, $m = 1, 2$. We used crude estimates of these parameters derived from the data using NONMEM; see also Moser et al[39]. Specifically, we used $\frac{1}{\sigma_1^2} = 3$, $\frac{1}{\sigma_2^2} = 43$, $\theta_1 = 42$, $\theta_2 = 0.5$, $\theta_3 = 0.655$, $\theta_4 = 3.34$, and $\theta_5 = 0.971$.

The resulting parameter estimates are given in Table 3.1, with corresponding fitted PK and PD curves shown in the left and right panels respectively of Figure 3.1. The prior and posterior standard deviations as given in the second and fourth columns are very different for θ_1, θ_3, θ_4 and θ_5, indicating that the data were able to inform these posteriors. In particular, the posterior for θ_3 reveals that the average maximal effect of erucic acid is fairly precisely estimated at 0.46 (95% BCI 0.40 to 0.51), and the estimated posterior median erucic level required to produce half of this maximal effect is only 0.84. These values are consistent with the fitted PD curve plotted in the right panel of Figure 3.1, where we replace all individual level parameters in (3.4) with their population level estimates. This plot also shows a scatterplot of the calculated $cssex$ versus Z_{ij} for each observation, a solid line showing the fitted model, and a dashed line depicting

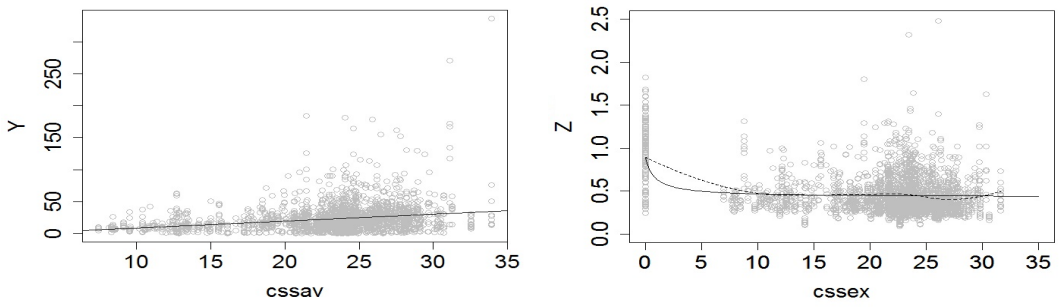

Figure 3.1: Left panel: plot of the PK data (Y_{ij} vs. estimated $cssav_{ij}$) along with the fitted regression line, which has slope 1.059; right panel: plot of the PD data (Z_{ij} vs. estimated $cssex_{ij}$) along with the fitted Emax curve.

a loess smooth of the data. The fit in both cases is far from perfect, in particular due to the missing information in the interval $0 < cssex < 7$. For larger $cssex$, the scatterplot has points that are dispersed far away from the fitted lines, perhaps indicating an increasing residual error variance with increasing $cssex$.

In order to see how sensible our PK estimates are, the left panel of Figure 3.1 offers an exploratory plot. If our estimates of the θ's are good, then according to (3.2), if we plot Y_{ij} vs $cssav_{ij}$, the resulting regression line should have slope approximately 1. Here $cssex_{ij}$ is calculated using our estimated $\hat{\theta}_1$, and $endo\_erucic_{ij}$ is the Y_{ij} value corresponding to $Dose = 0$. For individuals with no $Dose = 0$ measurement, we take $cssav_{ij} = cssex_{ij} + \hat{\theta}_2$, where $\hat{\theta}_2$ is the estimated posterior mean. The left panel of Figure 3.1 actually shows this plot without a few extremely outlying patients (discarding 69 of the 2558 observations), and also omitting observations with $dose = 0$. We obtain a fitted slope of 1.059, in agreement with our model. Thus, though our PK data are much less informative than our PD data, the Bayesian approach still permits refinement of our prior opinion.

3.3 Phase IIa Design

3.3.1 Approximate Linear PD Model

Having investigated the basic PK and PD properties of LO, we now proceed to design a basic Phase IIa trial to determine the toxicity and efficacy response of a few candidate dosing regimes. Consider n patients recruited in a clinical trial having a total duration of J biweekly periods ("bi-weeks"). We first record Z_{i0}, the observed C26 level for patient i at baseline, assuming that all participants have no exogenous erucic acid in their bodies at this time (perhaps after

a "washout" period) for $i = 1, ..., n$. At each bi-week $j = 1, ..., J$, we administer a daily dose of the drug to patient i that produces an erucic acid steady-state concentration of c_{ij}. We consider five values for c_{ij}: 0, 5, 10, 25 and 35. Hence when we say "a dose level of 5", we mean a dose large enough and given often enough to produce $cssex = 5$ for that particular patient. These c_{ij} levels were chosen based on the right panel of Figure 3.1, which indicates a meaningful reduction in Z for $cssex$ as low as 5. As the trial takes place, we observe Z_{ijk}, the C26 level for the i^{th} patient in the j^{th} period on the k^{th} day, $k = 1, ..., K_{ij}$ (where K_{ij} will typically be 14).

Next we define $X_{ij} = Z_{i0} - \sum_{k=1}^{K_{ij}} \frac{Z_{ijk}}{K_{ij}}$, the average *improvement* in C26 level for patient i in period j. For simplicity, in our initial trial design algorithm we assume $X_{ij}|\theta_{ij}, \sigma_x^2 \sim N(\gamma_{ij}, \sigma_x^2)$, and further assume

$$\gamma_{ij} = \mu_i^e + \beta_e c_{ij} \ , \tag{3.5}$$

where $\mu_i^e \sim N(\mu^e, \sigma_{\mu^e}^2)$ and the "e" superscript denotes efficacy. That is, for the moment we assume C26 improvement is *linear* in erucic acid concentration, with every person having their own baseline improvement parameter, μ_i^e. Adding the vague hyperprior distributions $\beta_e \sim Uniform(-1, 1)$ and $\sigma_x^2 \sim InverseGamma(0.1, 0.1)$ complete the efficacy model specification. Intuitively, if no dose is given, $E(X_{ij}) = \mu_i^e$ should be zero, since we are assuming Z_{i0} is the C26 level at baseline (pre-LO). Hence we assume the true values $\mu^e = 0$ and $\sigma_{\mu^e}^2 = 1$.

We assume toxicity outcomes to be binary variables (e.g., presence or absence of elevated platelet count in the blood, or bone marrow suppression) for each patient every bi-week, i.e., $W_{ij} = 1$ if toxicity was observed and 0 if no toxicity was observed. We assume that π_{ij}, the probability of observing a toxicity for patient i during bi-week j, is such that

$$logit(\pi_{ij}) = \mu_i^t + \beta_t c_{ij} \ , \tag{3.6}$$

where "t" denotes toxicity, and we now assume the linearity of the *logit* of the toxicity probability in c_{ij}. We assume $\mu_i^t \sim N(\mu^t, \sigma_{\mu^t}^2)$ and $\beta_t \sim Uniform(0, 1)$, so each individual has their own personal propensity for toxicity, and the drug cannot *reduce* toxicity. A simple calculation shows that $\mu_t = -4$ delivers $\pi_{ij} = 0.01$ at $c_{ij} = 0$. In what follows, we therefore assume the true values $\mu^t = -4$, and $\sigma_{\mu^t}^2 = 0.3$. We also experiment with $\mu^t = -2$ to check robustness of our method to this assumption.

Finally, to introduce correlation between the efficacy and toxicity responses, we can instead assume the following *joint* model for the subject level effects:

$$\begin{pmatrix} \mu_i^e \\ \mu_i^t \end{pmatrix} \sim N_2 \left(\begin{pmatrix} \mu^e \\ \mu^t \end{pmatrix} \ , \ \Sigma \right) \ , \tag{3.7}$$

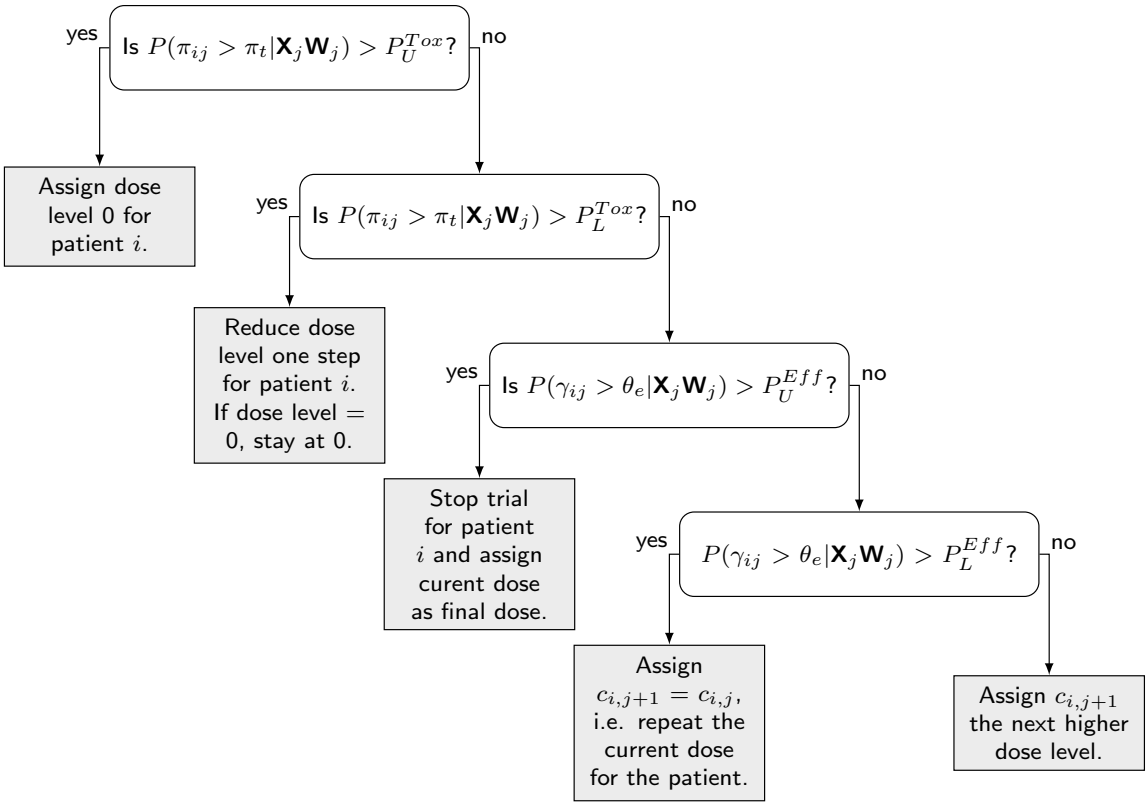

Figure 3.2: Dosing algorithm for Phase IIa trial design, patient i, bi-week $j \geq 4$.

where Σ takes the form $\begin{pmatrix} \sigma^2_{\mu^e} & \rho\sigma_{\mu^e}\sigma_{\mu^t} \\ \rho\sigma_{\mu^e}\sigma_{\mu^t} & \sigma^2_{\mu^t} \end{pmatrix}$. We put a modestly informative $Wishart(\Omega, 20)$ prior on Σ, and assume $\Omega = \left(\begin{smallmatrix} 100 & 0 \\ 0 & 0.02 \end{smallmatrix}\right)$ to preserve the imbalance in the scales of $\sigma^2_{\mu_e}$ and $\sigma^2_{\mu_t}$. Setting $\rho = 0.7$ and letting $\mathbf{X}_j$ and $\mathbf{W}_j$ denote all the efficacy and safety data, respectively, observed up to bi-week j, our dosing algorithm for the i^{th} patient at bi-week j is given in Figure 3.2. Note that our design is based entirely on posterior exceedance probabilities for toxicity (π_{ij}) and efficacy (γ_{ij}), where we use the thresholds $P_U^{Tox} = 0.8$, $P_L^{Tox} = 0.2$, $P_U^{Eff} = 0.9$, and $P_L^{Eff} = 0.4$. We selected threshold values of $\pi_t = 0.6$ for toxicity and $\theta_e = 0.8$ for efficacy based on clinical and practical considerations; below we check to make sure they deliver a design with acceptable operating characteristics.

To ensure adequate data before the trial's adaptive phase, we assign the starting dose ($c_{ij} = 5$) to all patients i for bi-weeks $j = 1, 2, 3$. Beginning with bi-week $j = 4$, we cycle through the flowchart in Figure 3.2 for each patient i and bi-week j, noting the dose level where each patient stopped due to high efficacy at the current dose level.

3.3.2 Emax PD Model

In this section we incorporate the Emax model we used in Subsection 3.2.2 into our Phase IIa design. In the design utilized in Section 3.3.1, we assumed an approximating linear PD model for the efficacy of Lorenzo's Oil. While this was computationally convenient for our design calculations, the right panel of Figure 3.1 suggests the relationship may be nonlinear. Therefore, in this section we return to model (3.3), estimated from the FAP data in Section 3.2.3. Since in our design we are estimating the difference of C26:0 from baseline, we replace (3.5) with

$$\gamma_{ij} = \frac{\theta_3 c_{ij}}{\theta_{4i} + c_{ij}} \ ,$$

where as before θ_3 and θ_{4i} correspond to the $Emax$ and $EC50$ parameters, respectively, and c_{ij} again denotes the steady state erucic acid concentration for subject i during bi-week j. By analogy with μ_i^e, each subject is allowed to have his or her own $EC50$ parameter. We adopt the priors $\theta_3 \sim Normal(\mu_3, \sigma^2_{\mu_3})$ and $\theta_{4i} \sim Normal(\mu_4, \sigma^2_{\mu_4})$, and also select the vague hyperprior distributions $\mu_3 \sim Uniform(-10, 10)$ and $\mu_4 \sim Uniform(-10, 10)$. We assign $\sigma^2_{\mu_3} = 1$ and $\sigma^2_{\mu_4} = 1$. Clearly θ_3 now drives the efficacy of LO, just as β_e did in our earlier design (3.5).

To begin with, we set $\beta_t = 0.15$ and assume no correlation between the efficacy and toxicity parameters, θ_{4i} and μ_i^t, although one could model such correlation using a bivariate normal prior as in (3.7). In what follows, toxicity is simply defined as in the previous subsection according to (3.6), where we assume $\mu_i^t \sim Normal(\mu^t, \sigma^2_{\mu^t})$ independently of the θ_{4i}, and where $\mu^t = -4$ and $\sigma^2_{\mu^t} = 0.3$. Our dosing algorithm remains the same as given in Figure 3.2, but where now $P_U^{Tox} = 0.9$ and $P_L^{Tox} = 0.5$. In our Emax model, the efficacy and toxicity models are assumed independent of each other, though the cut-offs have not changed (i.e., $\pi_t = 0.6$ and $\theta_e = 0.8$).

3.3.3 Phase IIa Emax Simulation Results

Our simulation results are based on a trial of $J = 12$ bi-weeks, various values of β_t, Σ as defined in the previous section, and use $L = 1000$ simulated datasets for each combination of n and β_e. For the ℓ^{th} simulated dataset, where $\ell = 1, ..., L$, we first generate X_{ij} and W_{ij} for all patients in each of the bi-weeks $j = 1, 2, 3$. Next, for bi-week $j = 4, ..., J$, we first generate X_{ij} and W_{ij} for the current bi-week and calculate the posterior estimates using $\boldsymbol{X}_j$ and $\boldsymbol{W}_j$ (i.e., all X_{im} and W_{im} for $m = 1, ..., j$). Based on the posterior estimates for bi-week j, we run through the dosing algorithm to determine the dose for each of the patients for bi-week $j + 1$. Finally at the end of the J^{th} bi-week, we obtain the posterior 95% equal-tail Bayesian credible interval (BCI)[8] for β_e. The power of the trial is estimated as the proportion of times among the L simulated datasets that the 95% BCI for β_e does not include 0 (since in our procedure this corresponds to rejecting $H_0 : \beta_e = 0$).

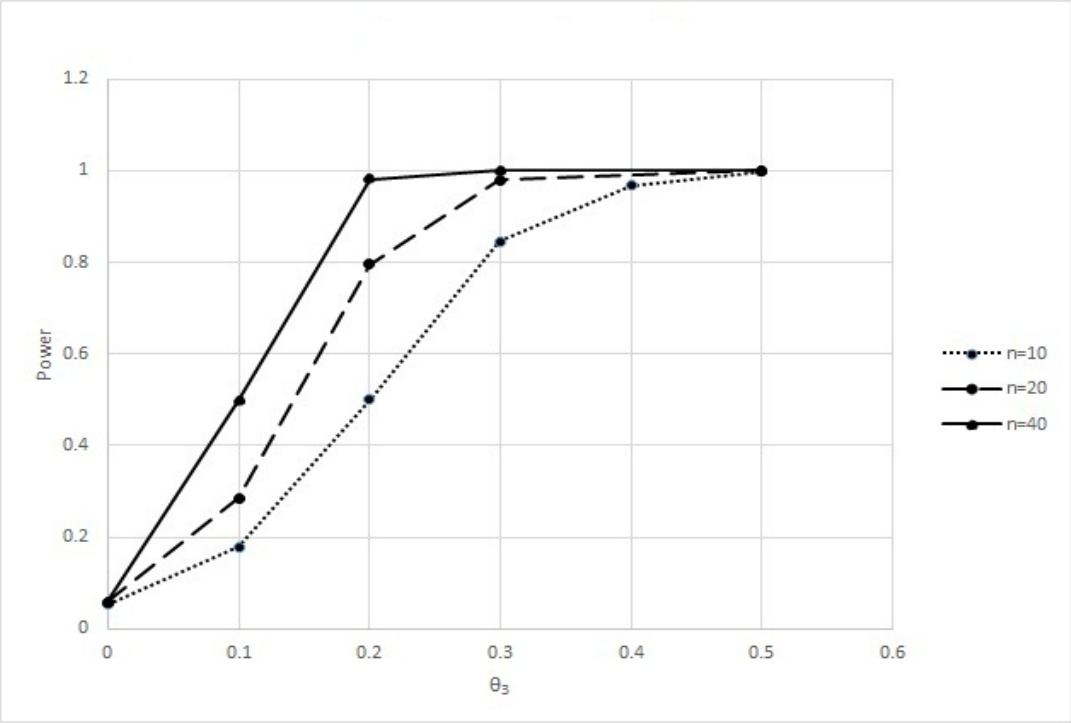

Figure 3.3: Plot depicting the power curve for the Emax model at each value of n with increasing θ_3 for $\theta_{4i} = 0.8931$ and $\mu^t = -4$ (low overall toxicity) and $\beta_t = 0.15$ (low drug toxicity).

For all subjects i, we assume $\theta_{4i} = 0.8931$, the estimate of θ_4 obtained in Section 3.2.3. We study the Type I error and power for varying values of n and θ_3, to give us an idea as to the operating characteristics of our design.

In spite of the nonlinearity introduced by the Emax model, trial operating characteristics remain reasonable. Table 3.2 provides some results under various choices of efficacy and toxicity parameters. The Type I errors (corresponding to $\theta_3 = 0$) are consistently around 0.05, and the power seems to escalate fairly quickly to desirable levels with increasing θ_3; see Figure 3.3 for power curves in the case where $\beta_t = 0.15$ (low drug toxicity). The θ_3 estimated in Section 3.2.3 is approximately 0.5, for which the power here is almost 1 even for n as small as 10. Similar simulations with the linear model also lead to reasonable operating characteristics slightly weaker than in the Emax model. Although the Emax model does perform well, it runs much more slowly than the linear model. Hence to the extent that the linear model can be thought of a good approximation to the true efficacy curve, opting for the linear model will lead to much faster computations. Certainly the right panel of Figure 3.1 offers little justification for *either* model, offering possibly more support for the simpler alternative.

As can be seen in Table 3.2, for $\theta_3 = 1$ cases shown when $\beta_t = 0.15$, the high assured efficacy of LO means more patients are responding very well and hence don't need to take a high dose (25 or 35). Also, on increasing the toxicity to $\beta_t = 0.3$ or $= 0.5$, we see fewer or even no patients being assigned to the higher doses, no doubt due in part to the flatness of the fitted Emax curve for larger LO doses.

3.4 Discussion

In Section 3.2 we successfully fit a hierarchical Bayesian model to obtain sensible PK/PD estimates for LO that can now be used to comprehend its clinical characteristics. We were then able to design an adaptive Phase IIa trial to evaluate LO while taking into account the effect of the treatment on each individual both in terms of efficacy and toxicity. In future this can serve as a safer trial for pediatric patients, and also produce reliable results. We were also able to base our Phase IIa model on the PK/PD of LO, which not only led to a more justifiable model but also can be used to update the PK/PD parameters during the course of the trial as more data are obtained to get a better understanding of LO. Further discussion and future work in this area is elaborated in Chapter 5.

n	β_t	θ_3	power	dose=0	dose=5	dose=15	dose=25	dose=35
10	0.15	0	0.049(0.007)	0.1	0.06	0.21	0.34	0.29
		0.2	0.503(0.016)	0.1	0.07	0.21	0.35	0.27
		0.5	0.998(0.001)	0.1	0.07	0.2	0.36	0.27
		1	1(0)	0.01	0.55	0.29	0.13	0.03
20		0	0.049(0.007)	0.1	0.07	0.23	0.32	0.28
		0.2	0.0792(0.009)	0.1	0.07	0.23	0.32	0.29
		0.5	1(0)	0.1	0.07	0.22	0.32	0.29
		1	1(0)	0	0.6	0.28	0.1	0.02
40		0	0.047(0.007)	0.09	0.06	0.25	0.32	0.28
		0.2	0.966(0.006)	0.09	0.07	0.24	0.32	0.28
		0.5	1(0)	0.09	0.06	0.25	0.31	0.28
		1	1(0)	0	0.71	0.23	0.05	0
10	0.3	0	0.072(0.008)	0.25	0.36	0.28	0.11	0
		0.5	0.998(0.001)	0.25	0.35	0.28	0.12	0
20		0	0.048(0.007)	0.27	0.34	0.27	0.12	0
		0.5	0.5(0.016)	0.28	0.34	0.25	0.12	0
40		0	0.055(0.007)	0.31	0.32	0.26	0.11	0
		0.5	1(0)	0.31	0.32	0.26	0.11	0
10	0.5	0	0.05(0.007)	0.48	0.16	0.35	0	0
		0.5	0.991(0.003)	0.47	0.17	0.36	0	0
20		0	0.054(0.007)	0.51	0.19	0.3	0	0
		0.5	1(0)	0.5	0.19	0.3	0	0
40		0	0.046(0.007)	0.5	0.19	0.31	0	0
		0.5	1(0)	0.5	0.2	0.3	0	0

Table 3.2: Estimated power (standard error) of the Phase IIa design (column 4) and proportion of patients being assigned to each dose (columns 5-9) using the Emax model for various choices of θ_3, β_t, and n at $\mu^t = -4$ and $\theta_{4i} = 0.8931$ for all i.

Chapter 4

PK/PD Data Extrapolation Models for Improved Pediatric Efficacy and Toxicity Estimation

4.1 Introduction and Motivating Dataset

As discussed in Section 1.2.3, secondary hyperparathyroidism (HPT) is a serious and frequently encountered problem in both pediatric and adult patients with chronic kidney disease (CKD) on dialysis. Secondary HPT is characterized by persistently elevated blood level of parathyroid hormone (PTH) levels caused by the release of excess amounts of PTH from enlarged parathyroid glands. The parathyroid glands are a component of a feedback loop that maintains blood calcium concentrations within a tight normal range[17]. Parathyroid hormone (PTH) secretion is increased in response to low calcium concentrations, and, alternatively, high calcium levels will turn off the secretion. The goals of treatment for secondary HPT in both children and adults include efforts to reduce the abnormally high concentrations of iPTH that characterize the disorder.

As discussed briefly in Section 2.3 cinacalcet is an oral calcimimetic agent that increases the sensitivity of the calcium-sensing receptor to extracellular calcium[41,42]. Cinacalcet is indicated for the treatment of secondary hyperparathyroidism (HPT) in adult patients with chronic kidney disease (CKD) on dialysis. It is not currently approved for use in children, though the drug maker, Amgen, is currently suing the FDA in an attempt to obtain six months of pediatric exclusivity for its use in children. Amgen is accusing the FDA of violating the Administrative Procedure Act by denying pediatric exclusivity, as well as by depriving the company of its Fifth Amendment right to due process[22]. Based on the understanding that the pathophysiology and progression of the disease, the role of the the calcium-sensing receptor, and the mechanism of action of

cinacalcet are similar in adults and children, the clinical rationale for using cinacalcet to lower elevated PTH concentrations in pediatric secondary HPT patients is considered similar to that of adult patients. Reduction of iPTH levels following treatment with cinacalcet may be associated with reduction in corrected calcium (cCa) levels, sometimes below the normal range, resulting in safety concerns.

In pediatric settings, it is extremely crucial that we are parsimonious with the patients recruited for experimentation, as pediatric settings pose myriad economical, logistical and ethical challenges [6,15,26]. The US federal government has taken action to promote research in these fields through various legislative actions. FDA has also promoted modernization of drug approvals through its Critical Path Initiative, and has issued guidance documents on extrapolating historical data of adults in pediatric trials for medical devices[12]. In particular, Bayesian statistical methods have been suggested as sensible ways to borrow information from adult trials, potentially allowing us to arrive at more reliable results even with relatively small pediatric datasets[5,8,30]. As seen above, various statistical methods exist for measured borrowing from existing historical datasets, including power priors[25] and commensurate priors[18]. However, the proper amount of information to borrow in any given situation (where the quality and commensurability of the adult information may be in doubt) remains difficult to determine.

In most regulatory settings, the usual approval process is preceded or accompanied by extensive studies to understand the drug's pharmacokinetic (PK) and pharmacodynamic (PD) properties. Yet in practice, although the PK/PD results are used in scale-up to support the determination of appropriate doses, these detailed data are not directly used in the confirmative trial setting. In this case, the PK/PD work is performed, checked, and then used to motivate subsequent studies of clinical safety and efficacy, but the degree of similarity between the PK/PD properties seen in children and adults is not explicitly used in the subsequent pediatric development program.

In this chapter, we attempt to utilize the rich PK/PD data to inform the amount of borrowing of information from adults during the later clinical phases of drug development. The basic idea here is that if the PK/PD properties of the drug are similar in adults and children, this helps justify borrowing more liberally from the adult data in clinical endpoints of the pediatric development program. PK/PD modeling is a well-understood and well-studied area in drug development, and FDA has also issued extensive guidance on these topics[13]. We will use hierarchical modeling of the clinical cinacalcet PK/PD data to quantitatively assess the similarity between adults and children, and use this information in various borrowing rules of clinical endpoints whose statistical properties can then be evaluated.

The remainder of this chapter unfolds as follows. In Section 4.2, we describe how we use

nonlinear mixed-effects (NLME) modeling to analyze the combined early pediatric and adult PK/PD data to determine the similarity of the PK/PD model parameters between the two groups. The NLME modeling (often referred to as a population PK/PD approach) is widely accepted in drug development as a tool to identify the pharmacostatistical characteristics of a pharmacological system. We then describe an approach for using measures of adult-child PK/PD similarity to inform a power prior model for combining subsequent adult and child data. Section 4.3 does this in the context of longitudinal efficacy data, while Section 4.4 extends our modeling to handle joint estimation of both efficacy and toxicity, again informed by the PK/PD results. Throughout we illustrate our approaches in the context of clinical trials of cinacalcet for treating pediatric secondary HPT, where we monitor both iPTH (efficacy) and cCa (safety) levels. After analyzing these data, we simulate the bias and mean squared error properties of our proposed procedure across a range of adult-child simulation scenarios in Section 4.5. Specifically, we simulate data with different levels of similarity between adults and children in each scenario. In Section 4.6 we offer an alternative model to describe the efficacy and toxicity of the drug using a piecewise model for the longitudinal data. We finally conclude the chapter with a brief discussion in Section 4.7.

4.2 PK/PD Modeling

We begin with a population-based NLME approach to develop a single PK/PD model that describes our observed concentration-time and effect-time data in pediatric and adult patients. The full model includes separate parameter estimates for adults and children, but can be reduced to a more parsimonious model where some parameters can be shared between the two groups. Our conceptual PK/PD model is pictured in Figure 4.1. The PK side (lower section) assumes linear PK and contains a dosing compartment with a first-order absorption process, a central compartment from which cinacalcet concentrations are measured that has first-order elimination, and a rapidly distributing peripheral compartment. The PD side (upper section) assumes a simple turnover model in which cCa is produced physiologically at a zero-order rate and is eliminated under a first-order process. These sections separately allow for drug concentrations to increase and decrease with each dose, and cCa to be at a baseline steady-state concentration when no cinacalcet is present. A novel semi-mechanistic allosteric activation model (to be the subject of a future manuscript by Amgen pharmacometric staff) accounts for the physiological feedback between iPTH and cCa. In this model, cCa production is stimulated by increases in iPTH from baseline, while iPTH production is directly inhibited by changes in CaR receptor occupancy by the innate ligand, ionized calcium, relative to the baseline (no cinacalcet treatment)

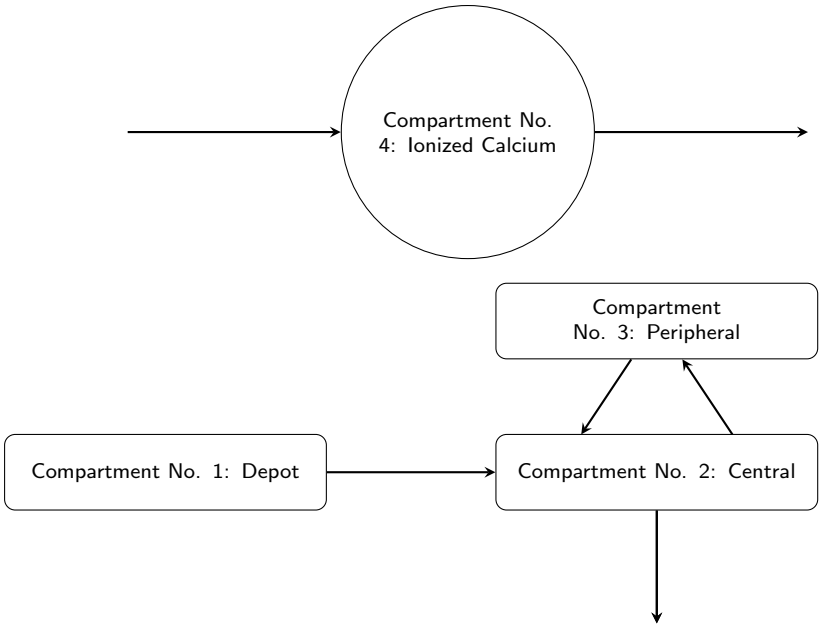

Figure 4.1: The four compartment structure assumed by the cinacalcet PK/PD model.

occupancy. The NLME modeling required is now fairly routinely implemented in software packages such as NONMEM (www.iconplc.com/innovation/nonmem/), which has become as standard in the pharmacological modeling community as BUGS has become for Bayesians.

In modeling PK-PD data, a choice must be made to analyze the PK and PD data simultaneously or sequentially[28,47,59,60,61]. Simultaneous modeling may be considered optimal from a hierarchical Bayesian point of view, but PD models are often highly complex (here, involving systems of differential equations), so even computing joint maximum likelihood estimates can be numerically difficult. Furthermore, we are more certain of the descriptive nature and precision of measurements in our PK models than our PD models, which means a simultaneous PK/PD approach risks the possibility of the PD side of the model misinforming our PK estimates. Hence, we chose to model these data sequentially. Specifically, the PK data were first modeled, and the empirical Bayes estimates of each individual's PK parameters were then used to generate predicted concentration-time profiles that together with the PD data help specify the PD parameter estimates.

In our setting, the PK/PD information consists of data from 7 clinical studies, 3 of which are adult studies as described in Section 1.2.3. Data from the aforementioned 7 studies were pooled and analyzed using NONMEM Version 7.3. Subjects who had no PK or PD concentrations available, or had screening failure even when PK or PD concentration or dosing records exist, were not included in the population PK/PD database. Any cinacalcet PK concentrations before the first dose administration were also excluded from the analysis, as were any cinacalcet PK or

PD concentrations below the quantification limit of the bioanalytical assay.

In this paper, our semimechanistic PD model features six key fixed effects θ_l, $l = 1, ..., 6$, with corresponding random effects η_{li} for the i^{th} patient, whose effect is modeled as $\theta_{li} = \theta_l e^{\eta_{li}}$. As previously stated, a novel three-compartment PK model was fit to the cinacalcet concentration data. This part of the analysis, the details of which will be the subject of a subsequent publication, provided empirical Bayes estimates of the individual PK parameters, which were then entered into the data set as fitted random effects. This produces patient-specific predicted cinacalcet concentrations that are then available to drive the PD response. The six PD parameters were initially permitted to have different values for adults and children (i.e., 12 parameters). Adding in the model's two variance parameters (for observation- and patient-level variability), we have a 14-parameter model that we refer to as Model 1. However this model suffered from poor NONMEM convergence.

Therefore, fitting Model 2 we set two of the six parameters to be the same for adults and children, resulting in a 12-parameter model. This model successfully converged with an objective function value (OFV, a goodness of fit metric similar to a sum of squares) not significantly different from that of Model 1 using a likelihood ratio test ($p > 0.05$ using the usual χ^2 statistic with df=2). However, when we set all six main effect parameters equal for adults and children (Model 3, an 8-parameter model), now the difference in fit between this model and Model 2 is statistically significant. Specifically, $OFV_{Model2} - OFV_{Model3}$ should follow a χ^2 distribution with $12 - 8 = 4$ degrees of freedom. For our data, this produces a p-value of $p = 0.002$. Thus, we have evidence that some, but not all of the PD model parameters can be safely assumed to be the same for adults and children, with a statistically significant gap between what might be thought of as "the full model" (Model 2, which converges and fits well) and "the null model" (Model 3, which erroneously assumes no differences at all between the adult and pediatric model fits). We emphasize that this p-value corresponds to a valid statistical test, but *not* one that the PD experiment was designed to run.

We suggest using some scale multiple of this p-value (say, Kp for some $K > 0$) as a sensible power (λ) in a power prior model for the clinical data. Specifically, K might be chosen to produce an adult EHSS that regulatory authorities do not find unacceptably large (say, many times larger than the actual pediatric sample sizes involved). Various EHSS definitions exist, but we can use definition (2.12) provided in Chapter 2 for a parameter of interest ξ namely

$$EHSS(\xi) = n_c \left[\frac{Var^{-1}(\xi|X^c, X^a, \lambda)}{Var^{-1}(\xi|X^c, \lambda)} - 1 \right], \tag{4.1}$$

which is the percent improvement in ξ's posterior precision (inverse variance) arising from using the chosen fraction of the adult data, expressed on the same scale as the child sample size.

Note that other metrics (e.g., $\sqrt{Var^{-1}}$) might have been chosen in (4.1), and that the answer will vary with the choice of ξ.

It may be useful to specify a target function to map the obtained p-value to an EHSS (or its ratio to the pediatric sample size n_c). These targets can be used by the drug developers as well as the regulatory authorities to specify how much adult information they are comfortable borrowing, and on what scale. We then run our hierarchical models for the longitudinal analysis with different degrees of adult data borrowing, and pick the one that results in a sufficiently close EHSS. The target function used in this chapter is

$$f(p) = \Psi \times p^{\alpha} \ , \ 0 < \alpha < 1. \tag{4.2}$$

In this function Ψ will dictate the scale for the range of the function and we may choose it to suit our interpretative need. For example, taking $\Psi = EHSS$ would relate p to the total effective number of adults borrowed, whereas $\Psi = EHSS/n_c$ would control the ratio of effective adults borrowed to the pediatric sample size. Figure 4.2 plots this function for a few sensible candidate values of α, all between 0 and 1 since the target should increase rapidly for smaller p and dampen as p becomes larger. The numerical scales here pertain to our cinacalcet clinical data sample sizes, which we describe in Section 3 below.

4.3 Hierarchical Modelling of Clinical Efficacy Data

Having measured the adult-child similarity in the PK/PD data, we turn to the longitudinal clinical data from adult and pediatric clinical efficacy studies of cinacalcet, in the context of a linear mixed effects Bayesian hierarchical model to study the drug's effect on iPTH. As discussed in Section 1.2.3, these data came from three adult Phase III clinical studies with subjects aged at least 18 years with secondary HPT receiving dialysis, and one pediatric Phase III study. Recapitulating, the adult studies have similar study designs and were randomized, placebo controlled clinical trials. Total sample size in these adult studies is N=1136. The pediatric data were collected from the 30-week, double-blind, placebo-controlled phase of the pediatric study. The study enrolled N=43 subjects aged 6 to 18 years with secondary HPT receiving dialysis. During the course of the studies, a cinacalcet dose was given daily, and labs including iPTH and cCa were measured every two weeks. In this chapter, we fully utilize the longitudinal information in the iPTH data collected during the whole study to evaluate the effect of cinacalcet in lowering iPTH level over time.

Many extrapolation techniques[10] have been suggested and used to streamline rare or pediatric drug development and thus speed approval of drugs for labeling. FDA also has issued guidance documents on borrowing of adult information for pediatric studies involving devices[12].

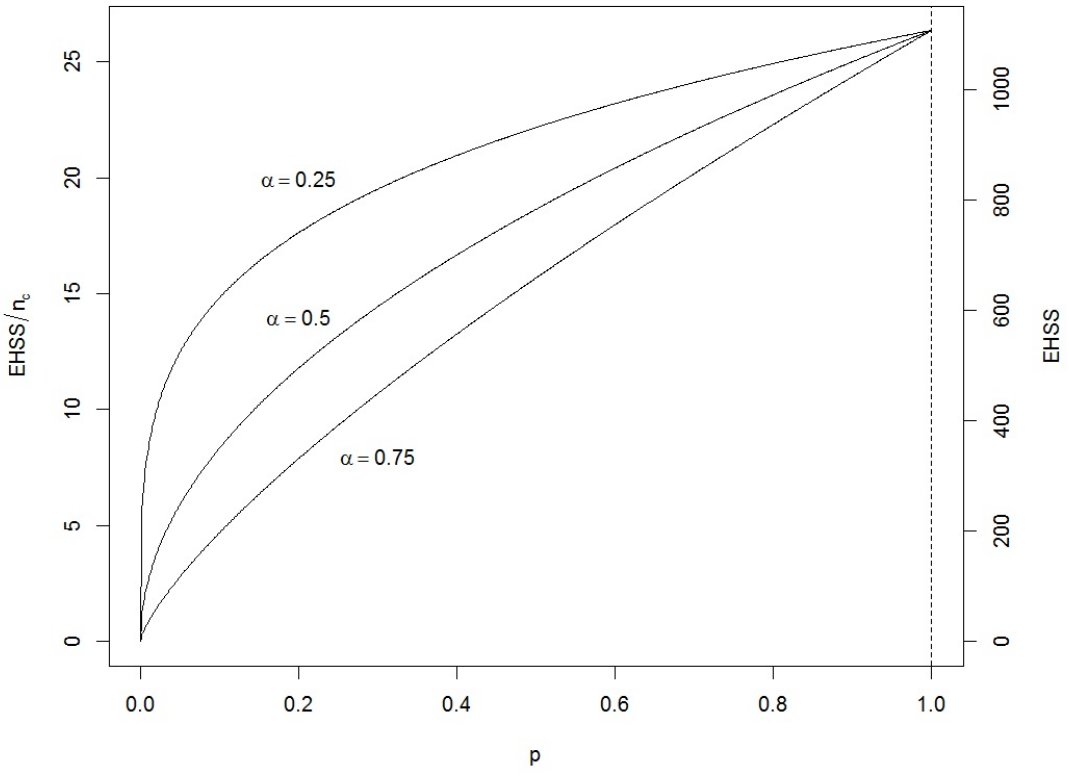

Figure 4.2: Plot of the target function $f(p)$ for various values of α. The right vertical axis corresponds to $\Psi = EHSS$ while the left vertical axis corresponds to $\Psi = EHSS/n_c$ using the sample sizes from our cinacalcet data.

However, borrowing information in the case of drugs is not as straightforward, in no small part because, unlike devices, drugs may well have PK and PD properties that differ between adults and children,[51] exactly the information we seek to incorporate in this chapter.

Based on the half-life of cinacalcet, a period of 7 days off treatment is expected to result in plasma levels of cinacalcet that do not have a pharmacodynamic effect. Thus data collected 7 or more days after the last dose date were excluded from the analyses. After this truncation of data, the number of patients with post baseline data in the pediatric dataset is $n_c = 42$, and $n_a = 1107$ in the adult studies. Let $X_{i,j}$ be the percent change from baseline in the iPTH level of patient i ($i = 1, ..., n_c, ..., (n_c + n_a)$) in the week of the patient's j^{th} observation, $j = 1, ..., m_i$ (where m_i varies from 3 to 25). That is:

$$X_{i,j} = \frac{ipth_{i,j} - \text{baseline ipth}_i}{\text{baseline ipth}_i} \times 100 \ .$$

This percentage change will be our outcome variable in the linear model. Let $t_{i,j}$ denote the week for the j^{th} observation on the i^{th} patient. This model for the pediatric group is

$$\begin{aligned} X^c_{i,j} &\sim Normal(\mu^{c,pth}_{i,j}, 1/\tau^c_e), \ i = 1, ..., n_c, \\ \text{where } \ \mu^{c,pth}_{i,j} &= \mu^c_{1i} t^c_{i,j} + I(drug^c_i = 1)(\mu^c_d t^c_{i,j}) \ . \end{aligned} \tag{4.3}$$

Our model specifies a linear relationship between the percent change from baseline in iPTH and time, where a negative slope indicates an effective treatment. Since by definition, the percent change from baseline is 0 at time point 0, there are no intercept terms in the pediatric or adult models for efficacy. Here, the μ^c_{1i} are subject-level random effects, assumed *a priori* to independently follow a Student's $t_3(\eta^c_0, \tau^c_\eta)$ specification, where η^c_0 is assigned a flat prior and τ^c_e is assigned a vague $Gamma(0.1, 0.1)$ prior. Since X^c_{ij} is a percent change of a quantity (iPTH) that takes only positive values, it cannot go below -100, but its upper limit is ∞. We expect a few extremely large outlying patient trajectories, justifying replacing the usual normal random effects distribution with the heavier-tailed t_3 to limit bias.

Similarly for the adults, we assume

$$\begin{aligned} X^a_{i,j} &\sim Normal(\mu^{a,pth}_{i,j}, 1/\tau^a_e), \ i = 1, ..., n_a, \\ \text{where } \ \mu^{a,pth}_{i,j} &= \mu^a_{1i} t^a_{i,j} + I(drug^a_i = 1)(\mu^a_d t^a_{i,j}) \ , \end{aligned}$$

where μ^a_{1i} are the subject-level random effects, assumed to independently follow a $t_3(\eta^a_0, \tau^a_\eta)$ specification. Similar to the child model, we assign a flat prior to η^a_0 and a vague $Gamma(0.1, 0.1)$ prior to τ^a_e.

As mentioned above, a natural approach for Bayesian borrowing of strength from an auxiliary dataset is through the power prior model, as described for instance by Ibrahim and Chen[25]. We

begin by assuming that $\eta_0^a = \eta_0^c = \eta_0$ and $\mu_d^a = \mu_d^c = \mu_d$. We then obtain the posterior $p(\mu_d|D_a, D_c)$ as proportional to $L(\mu_d|D_c)L(\mu_d|D_a)^\lambda\pi(\mu_d)$. Note that λ controls how much information will be borrowed from the auxiliary (adult) data to supplement the fully-utilized child data. As suggested by our earlier PD analysis in Section 4.2, since the p-value of the test comparing the "full"(adult and child parameters estimated separately) and "null" (adult and child parameters all identical) PD models provide a measure of the dissimilarity of the adult and child PD processes, we select $\lambda = Kp$ for some $K > 0$, where choice of K may be informed by our target function (4.2) and the resulting EHSS[19]. Under this approach that utilizes the p-value we allow our measure to be influenced by the sample size which could be a drawback hence we could use other approaches to arrive at our similarity measure such as those discussed in Section 4.7.

For fixed power priors, there is a one-to-one relationship between the power parameter and EHSS (though of course our MCMC-based estimates of the latter will be subject to error). The relationship is particularly straightforward in the normal likelihood setting; see for example Morita *et al.*[33,34] and Penello and Thompson[45]. In our case, we take the overall fitted slopes in the placebo group, η_0, and the drug group, $\eta_0 + \mu_d$, as our parameters of interest ξ.

In this chapter we use $\Psi = EHSS/n_c$ and $\alpha = 0.5$ in (4.2). Suppose we take $EHSS = n_a$, i.e., as shown in Figure 4.2, we are willing to borrow nearly all of the adult data's strength when p is close to 1 (no significant difference between adult and pediatric PD data).Then $\Psi = 1107/42 = 26.35$, and (4.2) gives $f(0.05) = 5.9$. This indicates a willingness on our part to base our conclusions on approximately six times as much adult data as pediatric data when the PK/PD datasets and models suggest no more than borderline statistically significant differences. At our actual obtained p value from the "null" and "full" model PD model comparison in Section 4.2, we have $f(0.002) = 1.17$. Hence we may seek a power λ that results in an EHSS of roughly $1.17 \times 42 = 49.5 \approx 50$ adult patients.

Table 4.1 provides our results, all computed using the `OpenBUGS` language[31]. We see that although using the p-value directly as the power (i.e., setting $K = 1$) does lead to EHSS values less than n_a, they are still very large compared to the pediatric data size n_c. By contrast, setting $K = 0.1$ (i.e., setting λ equal to 1/10 of the observed p-value) does appear to give more acceptable EHSS values (roughly equal and just a bit larger than n_c for our two parameters of interest, respectively). The table also gives results for the no-borrowing ($K = 0$) case for comparison. Of particular substantive interest here is that the overall slope in the drug group, $\mu^{pth} = \eta_0 + \mu_d$, is significantly (i.e., the 95% Bayesian credible interval excludes 0) negative in the high-borrowing ($K = 1$) and the more cautious ($K = 0.1$) cases, but not the no-borrowing ($K = 0$) case. Thus if we discard *all* of the adult data, the efficacy in children is not significant

		η_0					$\mu^{pth} = \eta_0 + \mu_d$				
K	λ	mean	sd	BCI (lower)	BCI (upper)	EHSS	mean	sd	BCI (lower)	BCI (upper)	EHSS
1	0.002	0.643	0.413	-0.079	1.557	207.198	-1.787	0.380	-2.475	-1.008	246.282
0.1	0.0002	1.015	0.708	-0.335	2.286	42.724	-1.622	0.634	-2.785	-0.285	61.477
0	0	1.254	1.006	-0.678	3.308	0	-1.574	0.995	-3.461	0.459	0

Table 4.1: Posterior estimates for the model coefficients using a power prior for varying values of $\lambda = Kp$. The three rows of the table correspond to $K = 1, 0.1,$ and 0.

statistically. Although $K = 0$ yielded a negative mean of slope for the drug group (showing some treatment effect), we arrived at a non-significant result possibly due to the small sample size of the pediatric study. This further emphasizes our motivation to extrapolate from adult data because it may have the potential to enable us to arrive at conclusive decisions as seen in this case.

4.4 Joint Clinical Efficacy and Toxicity Model

In order to jointly accommodate both efficacy and toxicity responses for each patient, we now specify a mixed effects model to simultaneously describe efficacy (measured by percent change in iPTH) and toxicity (measured by absolute drop in calcium level, cCa) of the drug among the adult and pediatric patients. We use the model described in the previous section for the efficacy measurements $X_{i,j}$. For the safety data, we now additionally define $Y_{i,j} = cCa_{i,j}$, the j^{th} corrected serum calcium observation for the i^{th} patient.

For pediatric patients (i.e., $i = 1, ..., n_c$), we now assume the bivariate normal likelihood $\begin{bmatrix} X_{i,j} \\ Y_{i,j} \end{bmatrix} \sim N_2(\boldsymbol{\mu}^c_{i,j}, \Sigma)$, where $\boldsymbol{\mu}^c_{i,j} = \begin{bmatrix} \mu^{c,pth}_{i,j} \\ \mu^{c,cCa}_{i,j} \end{bmatrix}$. Similarly we assume for the adults ($i = 1, ..., n_a$), $\begin{bmatrix} X_{i,j} \\ Y_{i,j} \end{bmatrix} \sim N_2(\boldsymbol{\mu}^a_{i,j}, \Sigma)$, where $\boldsymbol{\mu}^a_{i,j} = \begin{bmatrix} \mu^{a,pth}_{i,j} \\ \mu^{a,cCa}_{i,j} \end{bmatrix}$.

The data dispersion matrix is of the form $\Sigma = \begin{bmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_y^2 \end{bmatrix}$ and we assign priors such that $\frac{1}{\sigma_x^2}$ and $\frac{1}{\sigma_y^2}$ independently follow $Gamma(1, 1)$ distributions, and $\rho \sim Unif(-0.9, 0.9)$. Although theoretically the correlation can take values from -1 to 1, for numerical reasons we bound ρ away from these parameter space endpoints. Fitting a model similar to that in the univariate case above, we have

$$\begin{aligned}
\mu^{c,pth}_{i,j} &= \eta_0^{pth} t^c_{i,j} + I(drug^c_i = 1)\mu_d^{pth} t^c_{i,j} + \zeta_i^{c,pth} t_{i,j},\ i = 1, ..., n_c\ , \\
\mu^{c,cCa}_{i,j} &= \theta_c + \eta_0^{cCa} t^c_{i,j} + I(drug^c_i = 1)\mu_d^{cCa} t^c_{i,j} + \zeta_i^{c,cCa} t_{i,j},\ i = 1, ..., n_c\ , \\
\mu^{a,pth}_{i,j} &= \eta_0^{pth} t^a_{i,j} + I(drug^a_i = 1)\mu_d^{pth} t^a_{i,j} + \zeta_i^{a,pth} t_{i,j},\ i = 1, ..., n_a\ , \\
\text{and } \mu^{a,cCa}_{i,j} &= \theta_a + \eta_0^{cCa} t^a_{i,j} + I(drug^a_i = 1)\mu_d^{cCa} t^a_{i,j} + \zeta_i^{a,cCa} t_{i,j},\ i = 1, ..., n_a\ . \qquad (4.4)
\end{aligned}$$

We then assign $\begin{bmatrix} \zeta_i^{c,pth} \\ \zeta_i^{c,cCa} \end{bmatrix} \sim t_6(0, W_e)$ and $\begin{bmatrix} \zeta_i^{a,pth} \\ \zeta_i^{a,cCa} \end{bmatrix} \sim t_6(0, W_e)$. In order to choose the degrees of freedom for our bivariate t priors, we experimented with degrees of freedom 12, 6 and 3. Like the normal, the t_{12} prior failed to account for the extreme observations, while both the t_3 and t_6 were less affected. In what follows, we use the t_6 distribution, to allow ample flexibility to account for outliers, while limiting the degrees of freedom to ensure model identifiability (and hence MCMC convergence) is not compromised.

Note that we have introduced intercept terms θ_c and θ_a in the pediatric and adult models for toxicity, respectively. This is because, unlike in the case of iPTH where we are modeling a percent change from baseline (and hence at time point 0 we expect our percent change to be zero), in the case of toxicity we are modeling the absolute level of cCa, which is clearly nonzero at baseline. The random effects scale matrix, W_e is assigned an inverse Wishart prior with mean matrix W_0 and degrees of freedom w; in our implementation we let $W_0 = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ and set $w = 2$, the smallest value we can choose for this prior to remain proper. We then proceed to assign priors to parameters related to the cCa model analogous to those assigned for the iPTH model. Specifically, we assume $\eta_0^{cCa}, \mu_d^{cCa}, \eta_0^{pth}$, and μ_d^{pth} are assigned vague normal priors with mean 0, and the toxicity intercepts θ_a and θ_c are a priori $Uniform(0, 20)$.

We ran this model using the BUGS software (see code in appendix) and the results are given in Table 4.2. The results displayed for the iPTH (efficacy parameters) are similar to those obtained in the previous efficacy-only model. As in that model, of interest here is that the overall slope for efficacy in the drug group is once again more significant when we borrow more from the adults ($K = 1$ or $K = 0.1$ where it is borderline significant) than when we don't borrow ($K = 0$). The overall safety slope is also negative (as we would expect), but not quite significantly so even for $K = 1$. The mean posterior estimated ρ for all $\lambda's$ investigated lies near -0.15, which is practically negligible and suggests little correlation between the two endpoints.

We next compute, under $\lambda = 0.0002$, the posterior probability of the i^{th} pediatric patient having a mild toxicity (i.e., cCa value less than 8.4mg/dl) and a severe toxicity (cCa value less than 7.5mg/dl) at week 24, denoted by $p^{tox}_{24,8.4,i}$ and $p^{tox}_{24,7.5,i}$ respectively. Further, we also calculate a posterior probability of pediatric patients having good efficacy (defined as greater than 30% reduction in iPTH at week 24), denoted by $p^{eff}_{24,30,i}$, and excellent efficacy (defined as greater than 50% reduction in iPTH at week 24), denoted by $p^{eff}_{24,50,i}$. Figure 4.3 shows caterpillar plots corresponding to these four quantities from left to right. The horizontal lines in the plots are patient-specific 95% BCIs, the dots are patient specific point estimates, and the vertical lines in each plot denote the mean among that subgroup of patients. The first row contains the plots for pediatric patients in the drug group, and the second row shows pediatric patients in the placebo group.

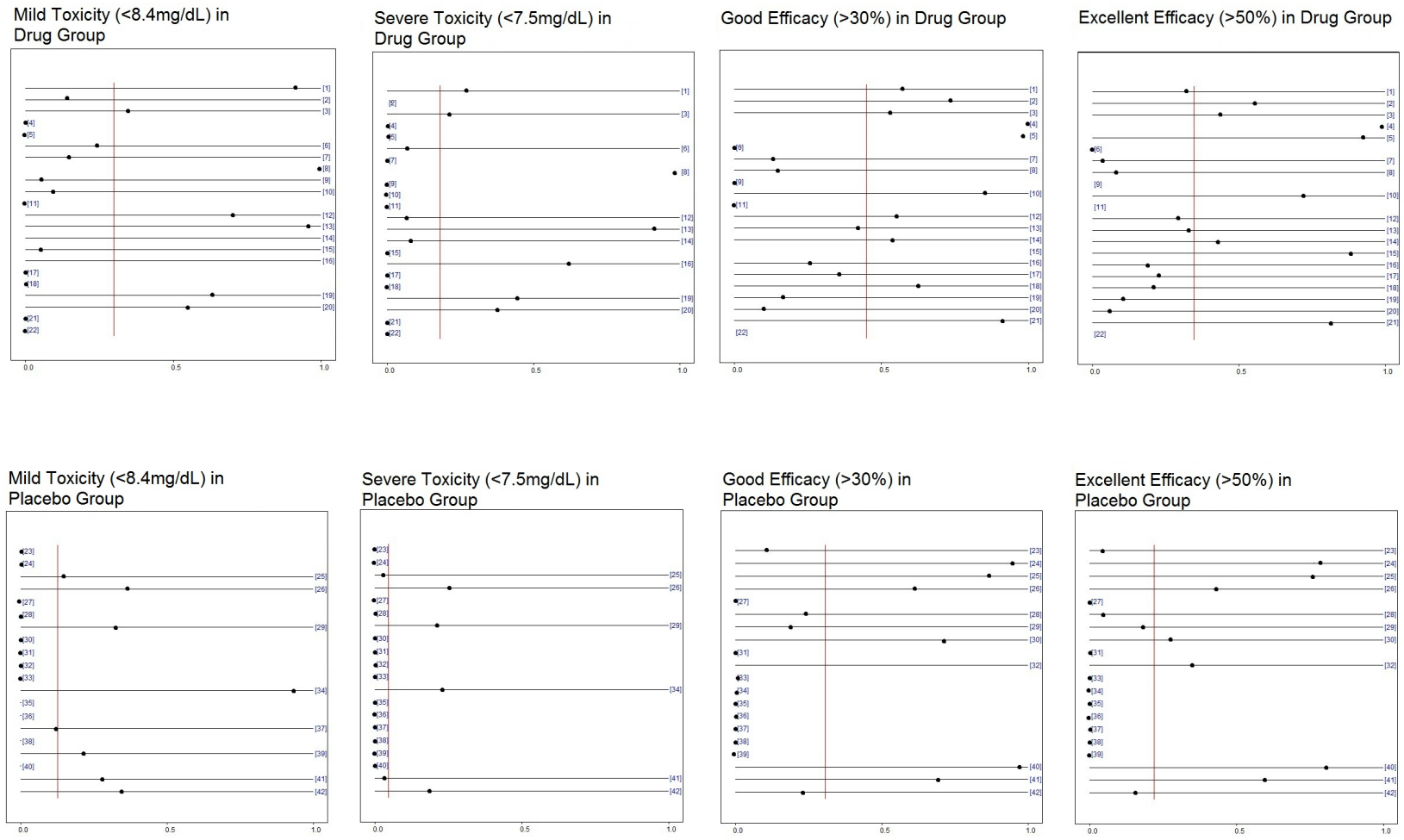

Figure 4.3: Caterpillar plots of the posterior probability of toxicity (cCa under 8.4 and under 7.5) and efficacy (more than 30% and 50% reduction in iPTH from baseline)at week 24 of each pediatric patient in the drug group when using power $\lambda = 0.0002$

		η_0^{pth}					$\mu^{pth} = \eta_0^{pth} + \mu_d^{pth}$				
K	λ	mean	sd	BCI (lower)	BCI (upper)	EHSS	mean	sd	BCI (lower)	BCI (upper)	EHSS
1	0.002	0.598	0.350	-0.048	1.399	438.612	-1.455	0.326	-2.082	-0.786	555.894
0.1	0.0002	1.077	0.789	-0.468	2.69	52.87	-1.324	0.715	-2.69	0.010	82.119
0	0	1.664	1.186	-0.594	4.063	0	-0.916	1.23	-3.257	1.592	0
		η_0^{cca}					$\mu^{cca} = \eta_0^{cca} + \mu_d^{cca}$				
K	λ	mean	sd	BCI (lower)	BCI (upper)	EHSS	mean	sd	BCI (lower)	BCI (upper)	EHSS
1	0.002	0.011	0.011	-0.012	0.032	588.796	-0.016	0.011	-0.037	0.006	596.968
0.1	0.0002	0.002	0.001	-0.037	0.041	162.565	-0.017	0.019	-0.054	0.019	170.748
0	0	-0.023	0.045	-0.114	0.063	0	-0.026	0.043	-0.111	0.057	0

Table 4.2: Posterior estimates for the joint model coefficients using a power prior for varying values of λ.

As can be seen, some patients have no posterior probability of dropping below the toxicity thresholds of 8.4 or 7.5, while Patient 8 is judged certain to have experienced either mild or severe toxicity. Specifically, 23 patients have a nonzero probability of mild toxicity, 15 in the drug group and 8 in the placebo group. From the caterpillar plot we can see that among the patients in the drug group, patients 8, 13, and 16 all have better-than-even probabilities of severe toxicity, and therefore should probably discontinue the treatment or seek dose adjustments or other interventions. Although according to the drug mechanism patients who see a higher improvement in iPTH are expected to have lower cCa, we don't see that in every individual patient, which is agreement with the posterior of ρ.

We do see more patients in the drug group who are more probable to show both moderate and high efficacy than in the placebo group, as can be seen in the shift in the vertical lines depicting the means of the groups. The average proportion of good and excellent efficacy is 0.45 and 0.35, respectively in the drug group, and 0.30 and 0.22, respectively in the placebo group. Thus the mean difference in response probability between the drug and placebo group is 0.15 and 0.13 for good and excellent efficacy, respectively. We do see the mean probabilities of mild and severe toxicity are also somewhat greater in the drug group than the placebo group: average mild and severe toxicity is 0.30 and 0.18, respectively in the drug group, and 0.13 and 0.05, respectively in the placebo group. Thus the shifts are 0.17 and 0.13 for mild and severe toxicity, respectively. Therefore we conclude the drug offers improved efficacy in reducing iPTH while increasing the risk of toxicity as measured as low cCa which may be subject to a risk/benefit evaluation.

4.5 Simulation Study

In this section we will simulate artificial datasets having characteristics based on the available cinacalcet data to evaluate our procedure's performance. We begin by simulating a dataset similar to our cinacalcet trial data data with $n_a = 200$ and $n_c = 10$ patients (so $n_a/n_c = 20$, similar to the real data), each with observations for every other week from week 1 to 29. The observations are the iPTH, X_{ij}, and cCa, Y_{ij}, and are simulated based on the models used to describe them in the previous sections with slight modifications for faster computations. Specifically, and as encouraged by the small ρ estimates in Section 4.4, we assume $\rho = 0$ and use independent univariate normal random effects distributions. We are able to drop the t_3 random effects distributions of Sections 4.3 and 4.4 since our simulated data should not feature outlying individuals to the extent that the real data did. As in Section 4.4, we generate $\begin{bmatrix} X_{ij}^c \\ Y_{ij}^c \end{bmatrix} \sim N_2(\boldsymbol{\mu_{ij}^c}, \boldsymbol{\Sigma})$, where again $\boldsymbol{\mu_{ij}^c} = \begin{bmatrix} \mu_{ij}^{c,pth} \\ \mu_{ij}^{c,cCa} \end{bmatrix}$ and are defined as noted previously but now for simplicity we assume $\zeta_i^{c,pth} \sim N(0, \tau_\zeta^{pth})$, $\zeta_i^{c,cCa} \sim N(0, \tau_\zeta^{cCa})$, and $\boldsymbol{\Sigma} = Diag(\sigma_X^2, \sigma_Y^2)$ (i.e., independence between efficacy and toxicity). Expressions for the adults follow similarly. We assign true values to the parameters (which help define the effect due to cinacalcet)$\eta_0^{pth,c} = 0, \eta_0^{cCa,c} = 0, \eta_0^{pth,a} = 0, \eta_0^{cCa,a} = 0, \theta_c = 10.5$, and $\theta_a = 10.5$. We also further fix the variances $\tau^{pth} = \sigma_X^2 = 66.67$, $\tau^{cCa} = \sigma_Y^2 = 0.013$, $\tau_\zeta^{pth} = 0.067$, and $\tau_\zeta^{cCa} = 0.0025$, values chosen to produce simulated data and random effects broadly similar to those in the cinacalcet data. We can then vary the true values of the $\mu_d^{pth,c}, \mu_d^{pth,a}, \mu_d^{cCa,c}$ and $\mu_d^{cCa,a}$ parameters (which help define the drug effect) to obtain different scenarios of similarity and dissimilarity between the adult and pediatric longitudinal trajectories.

In our set up we will begin by assigning equal values of $\mu_d^{pth,c} = -1.5$ and $\mu_d^{pth,a} = -1.5$. This is the case where the adults and pediatric patients are behaving similarly with respect to the drug efficacy. By contrast we also consider a case where borrowing is not sensible, namely, $\mu_d^{pth,c} = -1.5$ but $\mu_d^{pth,a} = 0$. A similar change could be made to the toxicity slope, either setting $\mu_d^{cCa,c} = \mu_d^{cCa,a} = -0.05$ or $\mu_d^{cCa,a} = -0.05$ but $\mu_d^{cCa,c} = -0.3$, the latter setting representing the worrisome case of much higher toxicity in children.

To evaluate our method, as shown in Table 4.3 we calculate the bias in the estimated slope corresponding to the drug effect. We have simulated data under 4 primary scenarios: when the adult and pediatric patients behave similarly both in terms of efficacy and toxicity, when the toxicity is the same but the two groups have different efficacy, when the efficacy in the two groups is the same but the toxicity is different, and when both the efficacy and toxicity are different in the two groups. Thus for the iPTH, $Bias_{ipth}$ at week 24 is the average of the difference between the posterior mean and true value, i.e., $((\hat{\eta}_{0,l}^{pth,c} + \hat{\mu}_{d,l}^{pth,c}) \times 24) - ((\eta_{0,l}^{pth,c} + \mu_{d,l}^{pth,c}) \times 24)$,

averaged over the $l = 1, ..., L = 500$ fake datasets. Similarly for the toxicity we define $Bias_{cCa}$ as the average difference between the true value and the posterior mean $(\hat{\eta}_{0,l}^{cCa,c} + \hat{\mu}_{d,l}^{cCa,c}) \times 24 - (\eta_{0,l}^{cCa,c} + \mu_{d,l}^{cCa,c}) \times 24$. We also calculate mean squared errors (MSE) at week 24, δ^{iPTH} and δ^{cCa}, given for instance by $\delta^{iPTH} = \frac{1}{L}\Sigma[((\hat{\eta}_{0,l}^{pth,c} + \hat{\mu}_{d,l}^{pth,c}) \times 24) - ((\eta_{0,l}^{pth,c} + \mu_{d,l}^{pth,c}) \times 24)]^2$. Intuitively these MSE and bias values should all be the same for $\lambda = 0$, as in that case we do not borrow anything from the adult dataset and thus their similarity or dissimilarity to the child data does not play a role. Otherwise we should have larger bias and MSE when the true values of the adult and child datasets are dissimilar.

Scenario		$\mu_d^{cca,a}=-0.05$, $\mu_d^{cca,c}=-0.05$,$\mu_d^{pth,a}=-1.5$ and $\mu_d^{pth,c}=-1.5$				
1	K	λ	$Bias_{CCa}$(SE)	MSE_{cCa}(SE)	$Bias_{iPTH}$(SE)	MSE_{iPTH}(SE)
	0	0	0.02(0.6)	0.36(0.56)	0.05(3.27)	10.67(15.08)
	0.1	0.0002	0.01(0.46)	0.21(0.3)	-0.03(3.07)	9.42(14.29)
	1	0.002	0.02(0.31)	0.1(0.15)	0.16(2.73)	7.46(11.99)
		$\mu_d^{cca,a}=-0.05$, $\mu_d^{cca,c}=-0.05$, $\mu_d^{pth,a}=0$ and $\mu_d^{pth,c}=-1.5$				
2	K	λ	$Bias_{cCa}$(SE)	MSE_{cCa}(SE)	$Bias_{iPTH}$(SE)	MSE_{iPTH}(SE)
	0	0	-0.01(0.59)	0.35(0.52)	-0.1(3.39)	11.47(17.94)
	0.1	0.0002	0.01(0.46)	0.21(0.33)	1.35(3.63)	14.98(53.56)
	1	0.002	-0.01(0.3)	0.09(0.13)	8.1(4.33)	84.4(82.8)
		$\mu_d^{cca,a}=-0.05$, $\mu_d^{cca,c}=-0.3$, $\mu_d^{pth,a}=-1.5$, and $\mu_d^{pth,c}=-1.5$				
3	K	λ	$Bias_{cCa}$(SE)	MSE_{cCa}(SE)	$Bias_{iPTH}$(SE)	MSE_{iPTH}(SE)
	0	0	0(0.57)	0.33(0.53)	-0.04(3.33)	11.04(18)
	0.1	0.0002	1.47(0.68)	2.61(2.95)	-0.01(3.29)	10.77(18.42)
	0.3	0.0006	2.47(0.69)	6.58(3.72)	0.08(3)	8.96(12.84)
	1	0.002	3.59(0.57)	13.21(4.18)	-0.01(2.73)	7.44(11.39)
		$\mu_d^{cca,a}=-0.05$, $\mu_d^{cca,c}=-0.3$, $\mu_d^{pth,a}=0$, and $\mu_d^{pth,c}=-1.5$				
4	K	λ	$Bias_{cCa}$(SE)	MSE_{cCa}(SE)	$Bias_{iPTH}$(SE)	MSE_{iPTH}(SE)
	0	0	0.01(0.55)	0.3(0.45)	-0.14(3.42)	11.68(18.08)
	0.1	0.0002	1.45(0.65)	2.52(2.34)	1.11(3.36)	12.49(23.06)
	0.3	0.0006	2.47(0.68)	6.56(3.67)	2.74(3.31)	18.44(27.54)
	1	0.002	3.6(0.56)	13.29(4.13)	8.07(4.38)	84.17(89.12)
		$\mu_d^{cca,a}=-0.05$, $\mu_d^{cca,c}=-0.3$, $\mu_d^{pth,a}=-1.5$, and $\mu_d^{pth,c}=0$				
5	K	λ	$Bias_{cCa}$(SE)	MSE_{cCa}(SE)	$Bias_{iPTH}$(SE)	MSE_{iPTH}(SE)
	0	0	0.01(0.68)	0.47(2.08)	-0.23(3.43)	11.77(22.66)
	0.1	0.0002	1.41(0.67)	2.42(2.71)	-1.25(3.28)	12.27(37.9)
	0.3	0.0006	2.5(0.71)	6.74(3.88)	-3.03(3.61)	22.15(52.65)
	1	0.002	3.57(0.54)	13.01(4.01)	-7.97(4.41)	83.04(101.8)

Table 4.3: Results from the simulation study assuming normal random effects.

As seen in Table 4.3, MSE_{iPTH} decreases and $Bias_{iPTH}$ remains roughly 0 as the power

multiplier K goes up in the case where the two datasets are similar (Scenario 1) because we are using more of the similar adult data for our pediatric estimation. However when the datasets are different, we see the bias and MSE for efficacy increasing in K, as we are now using data based on two different efficacy models. We can also see that the bias and errors for toxicity are little affected when the true values are different for only for efficacy which is due to our zero correlation assumption, and vice versa. For instance, the posteriors corresponding to toxicity are similar for Scenarios 1 and 2 (both good) and for Scenarios 3 and 4 (both bad), since only efficacy differs across these scenario pairs. Further, the bias and MSE in Scenarios 2, 3 and 4 increases as we borrow more from the discrepant adult data as expected. Incidentally we also checked a Scenario 5 where the pediatric patients did not benefit from the drug in terms of efficacy *and* were more toxic than the adults, and therefore were doing worse in both regards. The patterns for iPTH are similar to Scenarios 2 and 4 and for cCa were similar to Scenarios 3 and 4, as expected.

As a check on our work, we repeated the simulation reported in Table 4.3 now assuming t_3 random effects. To maintain the same random effect variances as in our normal simulation, we replaced τ_ζ^{pth} and τ_ζ^{cCa} by 3 times their previous values. Results shown in Table 4.4 show the same basic patterns.

To summarize our findings, $K = 0.1$ seems to offer nice improvements in MSE when borrowing is warranted, but limits the bias and keeps MSE from growing too large in the scenarios where borrowing is not warranted. We recommend investigators perform similar simulations in their own settings to help choose the crucial adult borrowing parameter K.

Scenario		$\mu_d^{cca,a}=-0.05$, $\mu_d^{cca,c}=-0.05$,$\mu_d^{pth,a}=-1.5$ and $\mu_d^{pth,c}=-1.5$				
1	K	λ	$Bias_{CCa}$(SE)	MSE_{cCa}(SE)	$Bias_{iPTH}$(SE)	MSE_{iPTH}(SE)
	0	0	-0.02(0.51)	0.26(0.54)	-0.24(2.83)	8.07(13.94)
	0.1	0.0002	0.02(0.35)	0.12(0.19)	0.05(2.57)	6.57(11.92)
	1	0.002	0(0.21)	0.05(0.06)	0.02(2.3)	5.26(8.45)
		$\mu_d^{cca,a}=-0.05$, $\mu_d^{cca,c}=-0.05$, $\mu_d^{pth,a}=0$ and $\mu_d^{pth,c}=-1.5$				
2	K	λ	$Bias_{cCa}$(SE)	MSE_{cCa}(SE)	$Bias_{iPTH}$(SE)	MSE_{iPTH}(SE)
	0	0	-0.03(0.52)	0.27(0.63)	0.09(2.68)	7.17(10.04)
	0.1	0.0002	-0.01(0.34)	0.12(0.19)	0.92(2.65)	7.85(12.23)
	1	0.002	0.01(0.22)	0.05(0.07)	7.23(5.88)	86.71(155.06)
		$\mu_d^{cca,a}=-0.05$, $\mu_d^{cca,c}=-0.3$, $\mu_d^{pth,a}=-1.5$, and $\mu_d^{pth,c}=-1.5$				
3	K	λ	$Bias_{cCa}$(SE)	MSE_{cCa}(SE)	$Bias_{iPTH}$(SE)	MSE_{iPTH}(SE)
	0	0	0.07(1.43)	2.06(41.17)	0.11(2.65)	7.01(12.79)
	0.1	0.0002	1.52(0.74)	2.85(3.07)	0.07(2.66)	7.05(11.82)
	1	0.002	4.81(0.57)	23.44(5.33)	0.21(2.23)	4.99(8.77)
		$\mu_d^{cca,a}=-0.05$, $\mu_d^{cca,c}=-0.3$, $\mu_d^{pth,a}=0$, and $\mu_d^{pth,c}=-1.5$				
4	K	λ	$Bias_{cCa}$(SE)	MSE_{cCa}(SE)	$Bias_{iPTH}$(SE)	MSE_{iPTH}(SE)
	0	0	-0.06(0.54)	0.3(1.09)	0.07(2.87)	8.25(18.53)
	0.1	0.0002	1.45(0.8)	2.76(3.5)	0.78(2.54)	7.03(10.79)
	1	0.002	4.79(0.63)	23.38(5.76)	7.37(5.7)	86.73(147.57)
		$\mu_d^{cca,a}=-0.05$, $\mu_d^{cca,c}=-0.3$, $\mu_d^{pth,a}=-1.5$, and $\mu_d^{pth,c}=0$				
5	K	λ	$Bias_{cCa}$(SE)	MSE_{cCa}(SE)	$Bias_{iPTH}$(SE)	MSE_{iPTH}(SE)
	0	0	0.05(0.81)	0.66(8.96)	0.01(2.99)	8.92(43.33)
	0.1	0.0002	1.51(0.79)	2.9(3.21)	-1.02(2.94)	9.64(16.77)
	1	0.002	4.73(0.64)	22.81(5.8)	-6.99(5.3)	76.84(127.97)

Table 4.4: Results from the simulation study assuming t_3 random effects.

4.6 Piecewise Linear Model for Efficacy

4.6.1 Efficacy-only Model

We next use the data from adult and pediatric clinical studies of Cinacalcet in the context of a *piecewise* linear mixed effects Bayesian hierarchical model to study the drug's effect on iPTH. From exploratory and other previous empirical and conceptual modeling, a knot, denoted by T_0, for the piecewise model is placed at week $t = 14$. Thus our longitudinal model for the children becomes:

$$\begin{aligned} X_{i,j} &\sim Normal(\mu^{pth}_{i,j}, 1/\tau_e),\ i = 1, ..., n_c \\ \mu^{pth}_{i,j} &= \eta_{0i} + \mu^c_1 t_{i,j} + \mu^c_2 (t_{i,j} - T_0) I(t_{i,j} - T_0 > 0) \\ &\quad + I(drug_i = 1)[\mu^{dc}_1 t_{i,j} + \mu^{dc}_2 (t_{i,j} - T_0) I(t_{i,j} - T_0 > 0)], \end{aligned}$$

where η_{0i} are the subject-level random effects, assumed to independently follow a $N(\eta_0, \tau_\eta)$ specification, and η_0 is assigned a vague flat prior. We assign a $Gamma(0.1, 0.1)$ prior on τ_e. Similarly for the adults, we assume:

$$\begin{aligned} X_{i,j} &\sim Normal(\mu^{pth}_{i,j}, 1/\tau_e),\ i = n_c + 1, ..., n_c + n_a \\ \mu^{pth}_{i,j} &= \eta_{0i} + \mu^a_1 t_{i,j} + \mu^a_2 (t_{i,j} - T_0) I(t_{i,j} - T_0 > 0) \\ &\quad + I(drug_i = 1)[\mu^{da}_1 t_{i,j} + \mu^{da}_2 (t_{i,j} - T_0) I(t_{i,j} - T_0 > 0)], \end{aligned}$$

where η_{0i} are again iid normal random effects centered around η_0, and all μ parameters and η_0 are assigned flat priors. Exploratory analyses suggest our use of a common data precision τ_e across both adults and children is reasonable. We assign commensurate priors on the coefficients for the slopes of the pediatric data in order to help their estimation borrow information from that in the model fit for adults. Specifically, we assume $\mu^c_1 \sim Normal(\mu^a_1, 1/\tau_\mu)$, $\mu^c_2 \sim Normal(\mu^a_2, 1/\tau_\mu)$, $\mu^{dc}_1 \sim Normal(\mu^{da}_1, 1/\tau_{\mu^d})$, and $\mu^{dc}_2 \sim Normal(\mu^{da}_2, 1/\tau_{\mu^d})$. Following Hobbs et al. (2011, 2012) and Section 2.3.2, the precisions τ_μ and τ_{μ^d} can be usefully assigned not the traditional gamma, but instead a *spike and slab* hyperprior. In our case, after a bit of experimentation with scale we selected:

$$\begin{aligned} \tau_\mu &\sim \begin{cases} Normal(20, 0.01) & \text{with probability } p; \\ Uniform(0.1, 5) I(0.1, 5) & \text{with probability } 1 - p\ , \end{cases} \\ \text{and } \tau_{\mu^d} &\sim \begin{cases} Normal(20, 0.01) & \text{with probability } p_d; \\ Uniform(0.1, 5) I(0.1, 5) & \text{with probability } 1 - p_d. \end{cases} \end{aligned} \tag{4.5}$$

Both these hyperpriors use a "spike" at 20, and a "slab" starting near the origin and going up to 5, but we permit different spike probabilities p and p_d for the baseline and the drug components. We take $p, p_d \overset{iid}{\sim} Bernoulli(q)$, where we experiment with various choice of q $(0.2, 0.5, 0.8)$.

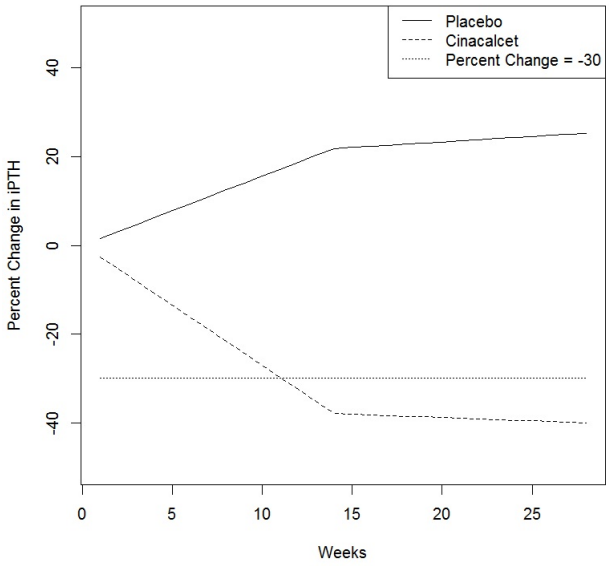

Figure 4.4: The piecewise linear model's regression line for the placebo and cincalcet group, for an Individual with intercept $\eta_{0,i}$ equal to zero

We ran this model in OpenBUGS, and the results are as shown in Tables 4.5- 4.6. As seen in Table 4.5, for both the adult and pediatric group, in the drug group there is a negative slope until the knot (pediatric CI = (-3.45,-1.98) and adult CI= (-3.31,-3.00)), followed by a slope insignificantly different from zero after the knot. On the other hand, the placebo group for both adults and pediatrics have a significantly positive slope until the knot and again a slope near to 0 after the knot. This indicates that the regression line as shown in Figure 4.4 for the drug initially drops steeply, and then becomes relatively flat after week 14. On the other hand, the line corresponding to the placebo group shows an initial increase in iPTH.

Figure 4.5 helps understand the spike and slab priors in (4.5). From these plots it looks like the prior and the posterior for the placebo group p are very similar, indicating that we are not learning much about adult vs. child commensurability from the adult population. On the other hand, the posterior for p_d shows that p_d is almost always 0 and hence the posterior is a draw from the Uniform distribution, suggesting little adult-child commensurability in the drug effect. This is also reflected in the plot for τ_{μ^d}, which is tightly located near 0.

According to our model, the mean posterior predictive probability of a greater than 30% reduction from baseline in iPTH was 0.049 (sd= 0.21) for placebo but 0.669 (sd 0.47) for Cinacalcet, indicating that the drug performs much better than placebo on this important clinical endpoint. Continuing with this theme, Table 4.6 shows the mean posterior probability

		Slope before the knot		Slope after the knot	
		Mean(SD)	CI	Mean(SD)	CI
Adult	Placebo	0.805(0.097)	(0.65,0.96)	0.04(0.109)	(-0.17,0.25)
	Drug	-3.155(0.093)	(-3.31,-3)	-0.026(0.102)	(-0.22,0.17)
Pediatric	Placebo	1.563(0.34)	(0.95,2.3)	0.247(0.381)	(-0.56,0.96)
	Drug	-2.712(0.373)	(-3.45,-1.98)	-0.151(0.542)	(-1.26,0.87)

Table 4.5: Posterior Mean Estimates of Slopes of Percent Change in Intact Hormone Over Time in Piecewise Linear Regression Model (Longitudinal Model) Using Commensurate Priors to Extrapolate Data from the Adult to the Pediatric Dataset

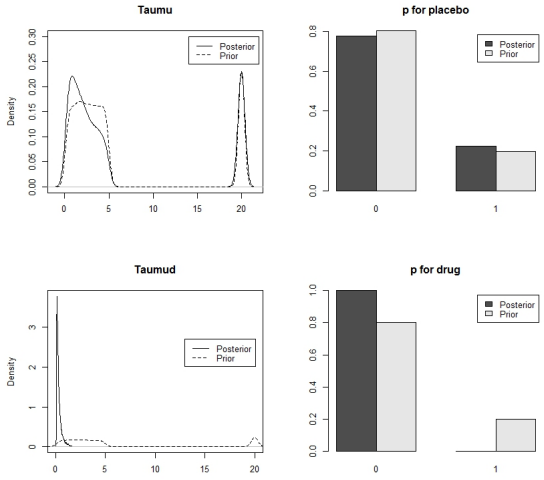

Figure 4.5: Plots Showing the Prior and Posterior Densities of the Spike and Slab Prior and The Corresponding p In Them

of true mean difference in percent change in iPTH for Cinacalcet compared with placebo at monthly time points being greater than π. The results again indicate very strong performance by the drug in all but the most extreme case (there is little chance of a 30% improvement in Week 2).

4.6.2 Joint Efficacy and Toxicity Longitudinal Model: iPTH and cca

To accommodate both efficacy and toxicity responses, we next tried our piecewise linear mixed effects model for both efficacy (as measured by iPTH) and toxicity (as measured by the excessive drop in the calcium levels, (cca) among the adult and pediatric patients. We use the model described above for efficacy measurements X_{ij} and now let $Y_{i,j} = cca_{i,j}$ where, $cca_{i,j}$ denotes the j^{th} observation on the corrected serum calcium value of the i^{th} patient. Let $Z_{i,j} = \begin{bmatrix} X_{i,j} \\ Y_{i,j} \end{bmatrix}$. Fitting a model similar to that in the iPTH model above, first for the children we have:

Week	2	4	6	8	10	12	13
π							
0	1(0.00)	1(0.00)	1(0.00)	1(0.00)	1(0.00)	1(0.00)	1(0.00)
10	0.999(0.02)	0.999(0.01)	1(0.00)	1(0.00)	1(0.00)	1(0.00)	1(0.00)
20	0.996(0.07)	1(0.02)	1(0.02)	1(0.01)	1(0.01)	1(0.01)	1(0.01)
30	0.035(0.185)	0.999(0.038)	0.999(0.025)	1(0.02)	1(0.018)	1(0.013)	1(0.015)

Table 4.6: Posterior Probability of True Mean Difference in Percent Change in Parathyroid Hormone for Cinacalcet Compared With Placebo at Monthly Time Points (Longitudinal Model) When Using Commensurate Priors to Extrapolate Information From Adult to Pediatric Data;

	Efficacy		Safety	
	Mean	SD	Mean	SD
Placebo	0.04	0.19	0.04	0.20
Cinacalcet	0.68	0.46	0.15	0.4

Table 4.7: Posterior predictive probability of $\geq 30\%$ reduction from baseline in iPTH and posterior predictive probability of $\leq 8.4\%$ in cCa.

$$Z_{i,j} \sim N_2(\boldsymbol{\mu}_{i,j}, W_e),\ i = 1, ..., n_c \text{ where } \boldsymbol{\mu}_{i,j} = \begin{bmatrix} \mu_{i,j}^{pth} \\ \mu_{i,j}^{cca} \end{bmatrix},$$

$$\mu_{i,j}^{cca} = \eta_{0_i}^{cca} + \mu_1^{cca,c} t_{i,j} + \mu_2^{cca,c}(t_{i,j} - T_0^{cca}) I(t_{i,j} - T_0^{cca} > 0)$$
$$+ I(drug_i = 1)[\mu_1^{cca,dc} t_{i,j} + \mu_2^{cca,dc}(t_{i,j} - T_0^{cca}) I(t_{i,j} - T_0^{cca} > 0)],$$

and μ_{ij}^{pth} is as given the previous subsection. Similarly for the adults we have $Z_{i,j} \sim N_2(\boldsymbol{\mu}_{i,j}, W_e)$,for $i = n_c + 1, ..., n_c + n_a$, with the expressions for μ_{ij}^{pth} and μ_{ij}^{cca} remaining unchanged. The 2X2 matrix W_e is assigned an Inverse Wishart prior with parameters W_0 and w where we encourage modest shrinkage by setting $W_0 = \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right]$ and $w = 20$. We then proceed to assign commensurate priors to parameters related to the CCA model analogous to those in the iPTH model; namely, $\mu_1^{cca,c} \sim Normal(\mu_1^{cca,a}, \tau_\mu^{cca})$, $\mu_2^{cca,c} \sim Normal(\mu_2^{cca,a}, \tau_\mu^{cca})$, $\mu_1^{cca,dc} \sim Normal(\mu_1^{cca,da}, \tau_\mu^{cca,d})$, and $\mu_2^{cca,dc} \sim Normal(\mu_2^{cca,da}, \tau_\mu^{cca,d})$. Finally τ_μ and $\tau_{\mu^{cca,d}}$ are also assigned the independent spike and slab priors given in (4.5).

We ran this model using the "Rjags" package in R, and the results are given in Tables 4.7 and 4.8. In Table 4.7, the results displayed related to iPTH are similar to those obtained in the previous subsection. On the safety side, cca levels below 8.4% are considered mildly toxic, so the probability of dropping below this (absolute) level is what we hope to control. Our results suggest a significant efficacy benefit for Cinacalcet (beating placebo 68% to 3%), but a corresponding slight increase in risk (15% to 4%).

Table 4.8 gives an entire picture of the joint regression model. The patterns found in the adult data are similar to those in the pediatric data. The slopes for the efficacy part are similar

Efficacy					
		Slope before the knot		Slope after the knot	
		Mean(sd)	CI	Mean(sd)	CI
Adult	Placebo	0.828(0.086)	(0.659,0.997)	-0.181(0.11)	(-0.398,0.036)
	Drug	-3.18(0.08)	(-0.929,0.293)	0.29 (0.10)	(0.093,0.487)
Pediatric	Placebo	1.660(0.31)	(1.100,2.3)	-0.17(0.37)	(-0.63,-0.86)
	Drug	-2.869(0.35)	(-3.61, -2.18)	0.016(0.54)	(-1.06, 1.03)
Toxicity					
		Slope before the knot		Slope after the knot	
		Mean(sd)	CI	Mean(sd)	CI
Adult	Placebo	0.003(0.001)	(0.001,0.005)	0.048(0.019)	(0.014,0.086)
	Drug	-0.088(0.002)	(-0.093,-0.084)	0.073(0.019)	(0.039,0.11)
Pediatric	Placebo	0.002(0.005)	(-0.009,0.014)	-0.024(0.037)	(-0.098, 0.044)
	Drug	-0.019(0.014)	(-0.049,0.009)	-0.045(0.035)	(-0.114, 0.024)

Table 4.8: Posterior mean estimates of slopes of percent change in iPTH over time and slopes for cCa levels in the efficacy-toxicity longitudinal model

to that obtained with the standalone efficacy model described in the previous subsection.

4.7 Discussion

In this chapter we have attempted to show how a careful study of PK/PD data can add valuable input into the controversial question of borrowing from adult data in subsequent pediatric drug evaluation. Looking at the PD model fit and posterior estimates may well offer a better understanding of the similarities in the effect of the drug between adults and children, adding to the justification for adult data borrowing. A quantitative analysis of the goodness of the PD model fit with or without separating the parameters by adult/peds population (full vs null models) and the posterior estimates can then sensibly inform the precise degree of adult clinical data borrowing. Although any "degree of similarity" measure will need to be scaled appropriately, our data analysis and limited simulation study have shed light on this important selection.

Our approach of using a p-value comparing the "full" and "null" PD models as the power in a second stage efficacy/toxicity model is just one way the PK/PD similarity information could be utilized. An alternate method to using the PD analysis results could be using the subject-level random effects η_{li} in the model and comparing their variance within and between groups. From NONMEM we can obtain the posterior mean for each individual corresponding to each of the fixed effect parameters, and carry out an ANOVA to obtain the variance within (V_w) and variance between (V_b) each age group. We then define a measure $\Upsilon_l = KV_{wl}/(V_{wl} + V_{bl})$ for each of the $l = 1, .., 5$ parameters, where again K is used to scale the result. Small values of Υ_l

suggest greater differences between adults and children. As such, the mean or median of these Υ_l can then be used as the power in the power prior method of analysis of the longitudinal data.

As yet another alternative, we might judge the similarity of the adult and pediatric results by directly comparing the posterior distributions of the corresponding adult- and child-specific main effects in the PD model. For example, we could measure the overlap of the areas under the two posterior densities. One way to do this would be through the integrated absolute difference, $\Delta = \int |P_c(\theta) - P_a(\theta)| d\theta$, where P_c and P_a denote the posterior distributions for parameter θ for the children and the adults, respectively. If the densities are entirely separate, the overlap will be zero and thus Δ should be 2. By contrast, if two densities are identical, we should obtain $\Delta = 0$. Hence, an alternate power to use in our power prior might be $K(1 - \Delta/2)$, where once again K is chosen to calibrate the overall amount of borrowing. Alternatives to Δ for judging posterior overlap include the Kullback-Liebler divergence, $\int P_c(\theta) \log \frac{P_c(\theta)}{P_a(\theta)} d\theta$, a discordance measure with many good properties and for which there exist `R` functions for ready computation.

The choice of *which* parameters to consider in order to evaluate the similarity between the groups remains open to debate. The complexity of the model (and smallness of the pediatric dataset) also means that our approximate posteriors may be rather unstable. It's also important to remember the difference between statistical and practical significance in this setting. Fitted posteriors may have little to no overlap between the adults and the children, but this may be due to a very large adult sample size, and not really connote practically meaningful differences between the groups. In such cases, limited borrowing from adult clinical data may still be warranted.

The target function $f(p)$ used in Figure 4.2 to map the p-value obtained from the PD analysis to the EHSS was attractively simple. However, we might also specify a piecewise function, which increases until a point (say, p=0.05 or 0.20) and then becomes flat, as another way to incorporate the restriction on maximum allowable EHSS. Also, we used a fat-tailed t distribution as our prior on the individual-level random effects, to avoid bias in the main effects due to extremely high iPTH percent change observations. This could also have been taken care of by modeling the data on the log scale and then fitting a similar model with the usual normal distributions on the random effects.

We also used simulated data to check that our procedure works as expected for a fixed power from the PD model. Although our results were consistent with what one would expect, FDA has historically discouraged borrowing of strength for safety data of any kind. A modification that can be used to address this issue would be to use differential borrowing, perhaps even taking $K = 0$ for toxicity but retaining $K > 0$ for efficacy. The method as described in the chapter

is well suited for EMA (EU regulatory) as it does not have similar restrictive opinion and in fact encourages extrapolation of safety endpoints. Also the joint approach may have broader application beyond model efficacy and toxicity. For example, it can evaluate co-primary efficacy endpoints that may or may not be highly correlated. In many disease areas, regulatory agencies have requested co-primary endpoints to demonstrate both histological improvement (treating the disease) and symptom improvement (treating the symptom).

As the drug is intuitively expected to cause a larger decrease in percent change of iPTH from baseline upto a point and then flatten, we also looked into the piecewise model to describe the efficacy and toxicity of the drug for the longitudinal model. The method to choose the knot remains an open question.

Chapter 5

Conclusions and Future Work

5.1 Summary Discussion

In Chapter 2, we began by describing binary and continuous longitudinal models used for Phase III adult and pediatric data. We used a hierarchical Bayesian framework implemented via power priors and commensurate priors, and showed they were effective ways to combine information both for of the binary Remicade data and for the longitudinal Cinacalcet data. Future work looks to larger simulation studies to compare these methods and study their properties. We will also be using estimated historical sample size as an added statistic to guide the amount of borrowing for the different methods, especially in the case of Cinacalcet where the current dataset is very small compared to the historical data.

In Chapter 3, we presented a fully Bayesian method for evaluating the PK and PD of LO, as well as a Bayesian adaptive design for a Phase IIa trial to determine a dose that shows improvement in biomarker response and is safe. Advantages of our method include the ability to get sensible estimates for a PK/PD model using what little data we have, possibly augmented by expert opinion. Our design is adaptive to interim individual toxicity and efficacy results, a desirable trait when dealing with a rare pediatric disease like X-ALD. The design also has a very reasonable Type I error (to guard against the chance that LO is not effective), as well as reasonably high power to detect the benefit of LO when it is effective, for all values of n that are feasible for our proposed Phase IIa study.

Limitations of our method here include failure of our preliminary PK/PD data to adequately inform the posterior estimate of θ_2, the parameter corresponding to the endogenous erucic concentration level. We have also not been able to incorporate individual-level errors in our model due to its already-high level of complexity. As for the Phase IIa design, our dose-finding algorithm can be rather sensitive to the precise level of toxicity assumed for LO (Table 3.2),

and these toxicities are somewhat vaguely defined at present. As such, our model may require modification if our assumptions are not consistent with real Phase IIa data.

Future work in this area aims to further develop our suite of adaptive clinical trial techniques. As discussed previously, this design is amenable to various possible models for improvement in biomarker response. The Emax model will often be preferred based on pharmacological justification for the drug's effect on the improvement in biomarker response. An advantage of using the Emax model in Phase IIa is we can update our model with the new data coming in; i.e., we can update our estimates of θ_4 and θ_3 in Phase IIa to make them more reliable. The subject pool for PK studies is often demographically or medically different from the subject pool in Phase IIa or Phase III, and hence being able to update the PK/PD parameters at a later stage may give us an estimate based on a more representative subject sample. Our ultimate goal is to design a Phase III study using CT/MR-detected brain abnormalities (instead of C26 response) as the primary endpoint.

The novel approach suggested in Chapter 4 was to use another auxiliary dataset, such as that from an earlier PK-PD analysis, to inform the degree of borrowing. Since the PK-PD process captures the drug's mechanism of action in terms of its clearance as well as the dose-response relationship[32], it seems like a sensible metric for estimating the similarity in the way the drug should affect adults and children. We quantitatively assessed auxiliary data similarity or dissimilarity, to come up with a more concrete justification for the degree of borrowing in such problems.

Future work in this area could further fine tune the procedure linking the PD analysis to the clinical data analysis. For simpler PD models, such as Emax models, we can set up models and analyze the PD and the clinical data simultaneously in `BUGS`. This will also allow us to obtain updated PD parameters to better understand the drug. We can also look at finding systematic methods to arrive at better scenarios for simulations. There are numerous other combinations of parameters besides the ones we selected that could have shed additional light on our method. A novel technique to deal with this could be to use fractional factorial designs to narrow down to a maximally informative subset of combinations from all possible simulation designs.

5.2 Zellner's g-prior for Rare and Pediatric Disease

In this section we will suggest the use of Zellner's g-prior for cautious strength borrowing in rare and pediatric disease as another possible area for future work. After a brief Subsection 5.2.1 review of the prior, in Subsection 5.2.2 we suggest a modified version to facilitate borrowing of information from one dataset to another. In Subsection 5.2.3 we propose applying this modified

Zellner's prior to simulated datasets to verify that the idea works before applying it to a dataset in Crohn's disease, using the modified g-prior in a logit model. Finally, we propose the use of propensity scores and other methods for choosing g in Subsection 5.2.4.

5.2.1 Review of Zellner's g-prior

Zellner's g prior for regression coefficients in a linear model was introduced by Goel and Zellner[16]. These authors begin by assuming $Y = X\beta + \epsilon$, a basic regression model. Zellner's informative g prior provides an intuitive way of determining how much the prior distribution on β should contribute to the posterior[14,16]. Zellner's prior is

$$\beta|\sigma^2, X \sim N_n(\beta_0, g\sigma^2(X^TX)^{-1})) \ .$$

Note that $\sigma^2(X^TX)^{-1}$ is nothing but $Var(\hat{\beta})$, where $\hat{\beta}$ is the usual maximum likelihood estimate of β. Zellner's g-prior incorporates the design matrix X in the covariance of the prior, making it similar to an empirical Bayesian technique to construct a default prior. This leads to a conditional posterior of the form

$$\beta|Y, \sigma^2, X \sim N\left(\frac{1}{g+1}(\beta_0 + g\hat{\beta}), \frac{g\sigma^2}{g+1}(X^TX)^{-1})\right) \ . \tag{5.1}$$

Zellner's informative prior produces a simple and intuitive interpretation of the relative weighting of the prior mean β_0 and the MLE $\hat{\beta}$, as well as helping from a computational perspective. Varying values of g provide different degrees of prior weighting according to (5.1). For instance, $g = 1$ implies a prior weight of 50%, $g = 10$ implies a prior weight of 10%, and $g \to \infty$ yields a diffuse prior.

5.2.2 Modified Zellner's g-prior

In this subsection we consider a modification to Zellner's g-prior. Consider two datasets, adult data D_0 and pediatric data D. Let X_0 denote the observed covariate values from the adult dataset, (say, age, the baseline Mayo score, duration of disease, etc.). Let the response variable in the adult data be Y_0, and in the pediatric data be Y. The covariates in the adult data are in columns of X_0 and those in the pediatric data are in the columns of X. Suppose we assume the model for the adult data is:

$$Y_0 \sim N_n(X_0\beta_0, \sigma_0^2 I) \ ,$$

and that we assign vague priors and compute the posterior estimates from the adult data as $\hat{\beta}_0$ and $\hat{\sigma}_0^2$. Then for the pediatric data, we can now assume the prior for β is

$$\beta \sim N(\hat{\beta}_0, g\hat{\sigma}_0^2(X_0^TX_0)^{-1}) \ .$$

This leads to another closed form posterior for β [30], though slightly more complicated than (5.1) namely, $P(\beta|Y,\sigma^2)$ is

$$N\left[\left(\frac{1}{\sigma^2}X^TX+\frac{1}{g\hat{\sigma}_0^2}X_0^TX_0\right)^{-1}\left(\frac{1}{\sigma^2}X^TY+\frac{1}{g\hat{\sigma}_0^2}X_0^TX_0\hat{\beta}_0\right),\left(\frac{1}{\sigma^2}X^TX+\frac{1}{g\hat{\sigma}_0^2}X_0^TX_0\right)^{-1}\right] \tag{5.2}$$

Thus we have modified Zellner's method by using a historical dataset. We have utilized estimates obtained from the historical dataset $\hat{\beta}$ and $\hat{\sigma}_0^2$ instead of using priors for them, which not only helps us avoid another set of possible vague priors but also are derived from the historical dataset, which is often much larger than the current dataset. This modification also uses the design matrix from a historical dataset instead of the current dataset, as Zellner's g prior did.

5.2.3 Simulated Datasets

In this subsection we will carry out some simulations to verify that the modified g-prior indeed works as we wish it to work. We start with an easy case. We simulate an adult dataset with two covariates each of size 100, and a pediatric dataset of size 50 also with two covariates each. We will simulate with various values of $\beta_{0,1}$, $\beta_{0,2}$, β_1 and β, all randomly generated numbers from a Uniform(-10,10). Also, we assume $x_{0,1,i} \sim N(45, sd = 17)$ and $x_{0,2,i} \sim N(6, sd = 2)$, $x_{1,i} \sim N(10, sd = 4)$ and $x_{2,i} \sim N(1.4, sd = 0.8)$, random numbers drawn from normal distributions centered around posterior means and data covariate values from our Remicade data set. We will study the effect of the prior by varying these parameters and confirming our ability to estimate the true parameter values. We then will illustrate the approach using our Remicade dataset.

5.2.4 Choice of g

According to the interpretation of g in the modified Zellner's g-prior, it is clear that when the two datasets are different, the choice of g will be crucial. Although g can be chosen based on expert opinion, we might also use our historical or adult data to help decide what g should be. This will be particularly helpful in cases when we do not have an expert's opinion to guide our choice of g, which is a problem that power prior and other methods also run into.

We propose to explore two methods in the following subsections. First we will consider a simple method of putting a hyperprior on g. If we have no other information, we can start with a vague hyperprior, such as an uniform with a large range. The second method we propose is

the use of a propensity score to estimate how similar the pediatric data are to the adult data, and then derive our g from that. This method, as we will see, comes with its own limitations and problems.

Hyperprior on g

Since g controls the amount of borrowing between datasets, its choice is critical. In Bayesian analysis, a vague prior is often assigned in information-poor settings. In our case, we can assign g a $Uniform(a, b)$ prior. Since g is always positive, we set $a = \epsilon > 0$ for $\epsilon = 0.1$. Meanwhile, b should be assigned a large value, which in our simulations was set to 150.

Under this prior, the conditional posterior of g can be derived as follows:

$$P(g|Y,\sigma^2) \propto \frac{1}{(b-a)g^{(\frac{n}{2})}} e^{-\frac{1}{2}[\frac{1}{\sigma^2}Y'Y + \frac{1}{g\hat{\sigma}_0^2}\hat{\beta_0}'X_0'X_0\hat{\beta_0} - A'\Sigma^{-1}A]} \left|\frac{1}{\sigma^2}X'X + \frac{1}{g\hat{\sigma}_0^2}X_0'X_0\right|^{-\frac{1}{2}} I_{a,b}(g),$$

where $A = \Sigma[\frac{1}{\sigma^2}X'Y + \frac{1}{g\hat{\sigma}_0^2}X_0'X_0\hat{\beta}_0]$ and $\Sigma = [\frac{1}{\sigma^2}X'X + \frac{1}{g\hat{\sigma}_0^2}\hat{\beta}_0'X_0'X_0\hat{\beta}_0]^{-1}$. This shows that the posterior of g does involve the data matrices, which indicates the opportunity to learn from the covariate data. Since this full conditional is not proportional to the prior, Bayesian learning seems possible, but its extent for any given problem remain unknown.

The results for the analysis of a few simulated datasets can be seen in Tables 5.1 and 5.2. Table 5.1 contains results from a setting where the two datasets were assumed to be from the same underlying distributions and models. Let us call this Sample 1. We simulated an adult dataset with two covariates of size 100 and a pediatric dataset of size 50 also with two covariates. True values of the parameters were $\beta_{0,1} = 4.01$ and $\beta_{0,2} = 8.311$ as well as the true $\beta_1 = 4.01$ and $\beta_2 = 8.311$, all of which are random numbers generated from a $Uniform(0, 10)$ distribution. Keeping both the underlying distributions the same, we next sampled $x_{0,1,i} \sim N(45, sd = 17)$, and $x_{0,2,i} \sim N(6, sd = 2)$, and for the pediatrics data we sampled $x_{1,i} \sim N(45, sd = 17)$ and $x_{2,i} \sim N(6, sd = 2)$. As mentioned before, we assign a prior $Uniform(0.1, 150)$ prior on g, as we cannot permit g to be identically 0. As can be seen from the results in Table 5.1, the posterior mean of g is only slightly smaller than the prior mean, which is approximately 75. This indicates minimal Bayesian learning about the g from the data.

We now consider a simulated dataset where the two sets are very different, which we call Sample 2. Again the simulated adult data have two covariates each of size 100, and the pediatric dataset is of size 50. We set the true $\beta_{0,1} = 0.461$ and $\beta_{0,2} = -6.296$ but the true $\beta_1 = 9.11$ and $\beta_2 = 3.22$. We further simulate from $x_{0,1,i} \sim N(45, sd = 17)$ and $x_{0,2,i} \sim N(6, sd = 2)$, and for the pediatric data $x_{1,i} \sim N(45, sd = 17)$ and $x_{2,i} \sim N(6, sd = 2)$. We again assign g

	Mean	SD	CI	
β_1	4.004	0.012	3.978	4.025
β_2	8.353	0.088	8.191	8.546
g	68.159	43.289	1.773	141.449

Table 5.1: The mean, standard deviations and confidence intervals for the model parameters and g in identical adult and pediatric datasets.

	Mean	SD	CI	
β_1	0.468	0.0668	0.332	0.429
β_2	-6.263	0.504	-7.313	-6.552
g	75.931	43.610	3.705	146.701

Table 5.2: The mean, standard deviations and confidence intervals for the model parameters and g in not similar adult and pediatric datasets.

the $Uniform(0.1, 150)$ hyperprior. The results can be seen in Table 5.2. The problem with the estimated g seems to persist. We do expect and get a larger value of g since the datasets are very different, however the posterior mean and standard deviation for g are still very similar to that of the prior, which again indicates weak learning from the data.

Thus we may conclude that assigning a vague prior on g doesn't lead to a significant learning about g, and hence doesn't work as a borrowing tool. If we do have some information on g, we can assign a better informed prior, such as a normal centered around a guess as to its value. However even in such cases, the results may again not learn much from the data, suggesting we either fix g or use some other method.

Use of Propensity Scores to Estimate g

Rosenbaum and Rubin [49] proposed the use of propensity scores as a method of adjusting in the presence of confounding by indication. Traditionally the propensity score is defined as an individual's probability of being treated with the intervention of interest given the complete set of all information about that individual. The propensity score provides a single metric that summaries all the information from explanatory variables[50]. Individual subjects may have the same or similar propensity scores, yet some will have received the intervention of interest and others will not.

In our case we redefine our propensity score as the probability that an individual with the given covariates *will belong to the pediatric dataset*. This means that if a patient has a higher propensity score, it is more likely that the patient has the same clinical characteristics as a child. In our case we will then calculate the scores for each adult patient in our dataset and obtain a

propensity score for each, using logistic regression in the usual fashion with adult or child status as the response. We could take the average of all the calculated adult propensity scores $\bar{p}$ to get one number for the dataset and use this number to decide if the two datasets are similar, or to guide data combination (say, by using the average score as a power in a power prior model).

We might also use this propensity score to arrive at our g. The average propensity score $\bar{p}$ lies between 0 and 1, and thus we need to convert it to our scale for g, any positive value. Our proposed approach is to find a function h that maps the propensity score to the g directly. One possible function to link the average propensity score $\bar{p}$ to g is $h(\bar{p}) = \frac{1}{\bar{p}}$. According to this function, $h(0) \to \infty$ and $h(1) = 1$ which is coherent with the properties of g. Note this approach assumes similarity of the covariate distribution is sufficient to justify borrowing strength from the adult data; the β coefficients related to the outcome measure may still differ between the adults and the children. As such the extent of borrowing here is not a function of the commensurability of the adult and pediatric data.

5.3 Conclusion

In conclusion, in this thesis we have seen some examples and methodologies where we use hierarchical Bayesian methods to facilitate data extrapolation that help us design and analyze clinical trials to tackle the limitations of current methods for pediatric trials or trials for rare disease treatments. The methods described and proposed in this thesis can indeed allow us to design trials that can be more sensitive toward patients, which is extremely crucial for such sensitive populations. These methods can also help us tie together various segments of the clinical trials and therefore base our analysis on more sound reasoning, as shown in Chapter 4. Therefore we have shown that using Bayesian modeling and specifically the methods proposed here, we can make data extrapolation a more reliable approach for superior clinical trials that both drug developers and regulatory are comfortable with and see profit in implementing.

References

[1] Asheuer M, Bieche I, Laurendeau I, Moser A, Hainque B, Vidaud M, Aubourg P. (2005). Decreased expression of ABCD4 and BG1 genes early in the pathogenesis of X-linked adrenoleukodystrophy. *Human Molecular Genetics* **14**:1293-1303.

[2] Aubourg P, Adamsbaum C, Lavallard-Rousseau F, Rocchiccioli F, Cartier N, Jambaqué I, Jakobezak C, Lemaitre A, Boureau F, Wolf C, and Bougneres PF. (1993). A two-year trial of oleic and erucic acids ("Lorenzo's oil") as treatment for adrenomyeloneuropathy. *New England Journal of Medicine*. **329**:745-752.

[3] Berger J , Pujol A, Aubourg P, Forss-Petter S. (2010). Current and future pharmacological treatment strategies in X-linked adrenoleukodystrophy. *Brain Pathology* **20**: 845-856

[4] Berger J, Molzer B, Fae I, Bernheimer, H. (1994). X-linked adrenoleukodystrophy (ALD): a novel mutation of the ALD gene in 6 members of a family presenting with 5 different phenotypes. *Biochemical and Biophysical Research Communications* **205**: 1638-1643

[5] Berry, SM, Carlin, BP, Lee, LL, and Muller, P. (2010). Bayesian Adaptive Methods for Clinical Trials. Boca Raton, Fl: Chapman and Hall/CRC Press.

[6] Best Pharmaceuticals for Children Act. (2002) Public Law 107-109. http://www.fda.gov/opacom/laws/pharmkids/pharmkids.html

[7] Caldwell Patrina HY, Murphy SB, Butow PN, Jonathan CC. (2004). Clinical Trials in Children. *The Lancet* **364.9436**: 803-811.

[8] Carlin BP, Louis TA. (2009) *Bayesian Methods for Data Analysis*, 3rd ed. Boca Raton, FL: Chapman and Hall/CRC Press.

[9] DiPiro JT, Spruill WJ, Wade WE, Blouin RA, Pruemer JM. (2010). *Concepts in Clinical Pharmacokinetics, 5th Edition*. Bethesda, MD: American Society of Health-System Pharmacists.

[10] Dunne, J, Rodriguez, WJ, Murphy, MD, Beasley BN, Burckart GJ, Filie JD, Lewis LL, Sachs HC, Sheridan PH, Starke P, Yao LP. (2011). Extrapolation of Adult Data and Other Data in Pediatric Drug-Development Programs. *Pediatrics*, **128**, e1242-1249

[11] Food and Drug Administration, Center for Devices and Radiological Health. (2010). *Guidance for the Use of Bayesian Statistics in Medical Device Clinical Trials*. U.S. Department of Health and Human Services, Rockville, MD. Released version of guidance document available online at `www.fda.gov/medicaldevices/deviceregulationandguidance/guidancedocuments/ucm071072.htm`

[12] Food and Drug Administration, Center for Devices and Radiological Health. (2016). *Leveraging Existing Clinical Data for Extrapolation to Pediatric Uses of Medical Devices: Guidance for Industry and Food and Drug Administration Staff*. U.S. Department of Health and Human Services, Rockville, MD. Released version of guidance document available online at `www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM444591.pdf`

[13] Food and Drug Administration, Center for Drug Evaluation and Research. (1999). *Guidance for Industry Population Pharmacokinetics* U.S. Department of Health and Human Services, Rockville, MD. Released version of guidance document available online at `https://www.fda.gov/downloads/drugs/guidances/UCM072137.pdf`

[14] Geinitz S. (2009), "Prior Covariance Choices and the g Prior", https://www.researchgate.net/publication/255665052

[15] Gerson J, Bright P, Lee CS, and Nelson RM. (2011). "Ethical Consideration in Conducting Pediatric Research." *Pediatric Clinical Pharmacology*. Springer Berlin Heidelberg. 219–244

[16] Goel, P, Zellner, A. (1986). "On Assessing Prior Distributions and Bayesian Regression Analysis with g Prior Distributions". Bayesian Inference and Decision Techniques: Essays in Honor of Bruno de Finetti. *Studies in Bayesian Econometrics 6*. New York: Elsevier. pp. 233-243.

[17] Goodman WG, Quarles LD. (2008). Development and progression of secondary hyperparathyroidism in chronic kidney disease: lessons from molecular genetics. *Kidney International*. **74**:276–288.

[18] Hobbs BP, Carlin BP, Mandrekar SJ, and Sargent DJ. (2011). Hierarchical commensurate and power prior models for adaptive incorporation of historical information in clinical trials. *Biometrics* **67**:1047–1056.

[19] Hobbs BP, Carlin BP, and Sargent DJ. (2013). Adaptive adjustment of the randomization ratio using historical control data. *Clinical Trials*. **10(3)**:430–440.

[20] Hobbs BP, Sargent DJ, and Carlin BP. (2012). Commensurate priors for incorporating historical information in clinical trials using general and generalized linear models. *Bayesian Analysis* **7(3)**:639–674.

[21] http://www.sensipar.com/about-sensipar/how-sensipar-works.html

[22] http://www.genengnews.com/gen-news-highlights/amgen-sues-to-reverse-fda-denial-of-pediatric-exclusivity-for-sensipar/81254428

[23] Hyams J, Damaraju L, Blank M, Johanns J, Guzzo C, Winter HS, Kugathasan S, Cohen S, Markowitz J, Escher JC, Veereman-Wauters G, Crandall W, Baldassano R, Griffiths A. (2012). Induction and maintenance therapy with infliximab for children with moderate to severe ulcerative colitis. *Clinical Gastroenterology and Hepatology*, **10(4)**:391-399.

[24] Hyams JS, Lerer T, Griffiths A, Pfefferkorn M, Stephens M, Evans J, Otley A, Carvalho R, Mack D, Bousvaros A, Rosh J, Grossman A, Tomer G, Kay M, Crandall W, Oliva-Hemker M, Keljo D, LeLeiko N, Markowitz J. (2010). Outcome following infliximab therapy in children with ulcerative colitis. *The American Journal of Gastroenterology*, **105(6)**:1430-1436.

[25] Ibrahim JG and Chen MH. (2000). Power prior distributions for regression models. *Statistical Science*, **15**:46–60.

[26] Institute of Medicine (US) Forum on Drug Discovery, Development, and Translation. (2008). Addressing the Barriers to Pediatric Drug Development: Workshop Summary. Washington (DC): National Academics Press (US). Forum on Drug Discovery, Development, and Translation.

[27] Koehler W, Sokolowski P. (1999). A new disease-specific scoring system for adult phenotypes of X-linked adrenoleukodystrophy. *Journal of Molecular Neuroscience* **13(3)**: 247-252.

[28] Lacroix BD, Friberg LE, and Karlsson MO. (2012). Evaluation of IPPSE, an alternative method for sequential population PKPD analysis. *Journal of Pharmacokinetics and Pharmacodynamics* **39**:177–193.

[29] Lewis JD, Chuai S, Nessel L, Lichtenstein GR, Aberra FN, Ellenberg JH. (2008). Use of the Non-invasive Components of the Mayo Score to Assess Clinical Response in Ulcerative Colitis. *Inflammatory Bowel Diseases*. **14(12)**:1660-1666. doi:10.1002/ibd.20520.

[30] Lindley DV and Smith AFM. (1972). Bayes estimate for the linear model. *Journal of the Royal Statistical Society. Series B (Methodological)*, **34**, 1–41.

[31] Lunn D, Jackson C, Best N, Thomas A, and Spiegelhalter DJ. (2012). *The BUGS Book: A Practical Introduction to Bayesian Analysis*. Boca Raton, FL: Chapman and Hall/CRC Press.

[32] Macdougall J. (2006). *Analysis of Dose-Response Studies-Emax Model*. New York: Springer, pp. 127-145.

[33] Morita S, Thall PF, and Müller P. (2008). Determining the effective sample size of a parametric prior. *Biometrics*, **64**, 595–602.

[34] Morita S, Thall PF, and Müller P. (2012). Prior effective sample size in conditionally independent hierarchical models. *Bayesian Analysis*, **7**, 591–614.

[35] Moser J, Douar AM, Sarde CO, Kioschis P, FEIL R, MOSER H, POUSTKA A, MANDEL JL, AUBOURG P. (1993). Putative X-linked adrenoleukodystrophy gene shares unexpected homology with ABC transporters. *Nature*, **361**: 726-730

[36] Moser AB, Kreiter N, Bezman L, Lu S, Raymond GV, Naidu S, Moser HW. (1999). Plasma very long chain fatty acids in 3,000 peroxisome disease patients and 29,000 controls. *Annals of Neurology*, **45**(1):100-110.

[37] Moser HW, Raymond GV, Dubey P. (2005). Adrenoleukodystrophy: new approaches to a neurodegenerative disease. *Journal of American Medical Association* **294**: 3131-3134.

[38] Moser HW, Raymond GV, Koehler W, Sokolowski P, Hanefeld F, Korenke GC, Green A, Loes DJ, Hunneman DH, Jones RO, Lu SE, Uziel G, Giros ML, and Roels F. (2003). Evaluation of the preventive effect of glyceryl trioleate-trierucate ("Lorenzo's oil") therapy in X-linked adrenoleukodystrophy: results of two concurrent trials. In *Peroxisomal Disorders and Regulation of Genes*, New York: Springer, pp.369-387.

[39] Moser HW, Raymond GV, Lu SE, Muenz LR, Moser AB, Xu J, Jones RO, Loes DJ, Melhem ER, Dubey P, Bezman L, Brereton NH, Odone A. (2005). Follow-up of 89 asymptomatic patients with adrenoleukodystrophy treated with Lorenzo's oil. *Journal American Medical Association* **62(7)**:1073-1080.

[40] Moser HW, Smith KD, Watkins PA, Powers J, Moser AB. (2001). X-linked adrenoleukodystrophy. In *The metabolic and molecular bases of inherited disease* (8th ed.), CR Scriver and AL Beaudet, eds., New York: McGraw-Hill, pp. 3257-3301.

[41] Nemeth EF and Bennett SA. (1998). Tricking the parathyroid gland with novel calcimimetic agents. *Nephrology Dialysis Transplant*. **13**:1923–1925.

[42] Nemeth EF, Heaton WH, Miller M, Fox J, and Balandrin MF, Van W, Bradford C, Colloton M and Karbon W, Scherrer J, Shatzen E, Rishton G, Scully S, Qi M, Harris R, Lacey D and Martin D . (2004). Pharmacodynamics of the type II calcimimetic compound cinacalcet HCl. *Journal of Pharmacology and Experimental Therapeutics*. **308**:627–635.

[43] Orphan Drug Act. (2013) Public Law 97–414. http://www.fda.gov/ForIndustry/DevelopingProductsforRareDiseasesConditions/HowtoapplyforOrphanProductDesignation/ucm364750.htm

[44] Pediatric Research Equity Act. (2003). Public Law 108–155. http://www.gpo.gov/fdsys/pkg/PLAW-108publ155/html/PLAW-108publ155.htm

[45] Pennello G and Thompson L. (2016). Design Considerations for Bayesian Clinical Studies: Prior Effective Sample Size and Type 1 Error Level. Proceedings of the ASA Biopharmaceutical Section FDA-Industry Statistics Workshop, Washington DC.

[46] Pennello G and Thompson L. (2008). Experience with Reviewing Bayesian Medical Device Trials. *Journal of Biopharmaceutical Statistics*, **18** 81–115.

[47] Proosta JH, Schierea S, Elevelda DJ and Wierda JMKH. (2007). Simultaneous versus Sequential Pharmacokinetic–pharmacodynamic Population Analysis using an Iterative Two–Stage Bayesian Technique. *Biopharmaceutics and Drug Disposition*, **28**: 455–473.

[48] Rizzo WB, Leshner RT, Odone A, Dammann AL, Craft DA, Jensen ME, Jennings SS, Davis S, Jaitly R, and Sgro JA. (1989). Dietary erucic acid therapy for X-linked adrenoleukodystrophy. *Neurology* **39**: 1415-1422.

[49] Rosenbaum PR, Rubin DB. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* ;**70**:41-55.

[50] Rosenbaum, P, Rubin, D. (1984). Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. Journal of the American Statistical Association, **79**(387), 516-524. doi:10.2307/2288398

[51] Selevan SG, Kimmel CA, and Mendola P. (2000). Identifying Critical Windows of Exposure *Children's Health. Environ Health perspect*, **108**(Suppl 3):54

[52] Singh I, Pujol A. (2010). Pathomechanisms underlying X-adrenoleukodystrophy: a three hit hypothesis. *Brain Pathology*. **20**(4): 838-844.

[53] Thompson L, and Pennello G. (2015). Borrowing from adult data to make inferences about the effectiveness of medical devices in a pediatric population. Technical report, Food and Drug Administration, Center for Devices and Radiological Health.

[54] Uchizono JA, Lane JR. (2006). *Empirical Pharmacokinetic/Pharmacodynamic Models*. New York: John Wiley & Sons, Inc. http://dx.doi.org/10.1002/9780470087978.ch20, Pharmacometrics.

[55] Upton RN, Mould DR. (2014). Basic Concepts in Population Modeling, Simulation, and Model-Based Drug Development: Part 3-Introduction to Pharmacodynamic Modeling Methods. *CPT: Pharmacometrics & Systems Pharmacology* online. American Society for Clinical Pharmacology and Therapeutics.

[56] van Geel BM, Assies J, Haverkort EB, Koelman JH, Verbeeten B, Jr, Wanders RJ, Barth PG. (1999). Progression of abnormalities in adrenomyeloneuropathy and neurologically asymptomatic X-linked adrenoleukodystrophy despite treatment with "Lorenzo's oil". *Journal of Neurology, Neurosurgery, and Psychiatry* **67**:290-299.

[57] Sensipar. http://www.sensipar.com/about-sensipar/how-sensipar-works.html

[58] Viele K. (2016). Issues in the incorporation of historical data in clinical trials. Proceedings of the ASA Biopharmaceutical Conference, Washington DC.

[59] Zhang L, Beal SL, and Sheiner LB. (2003). Simultaneous vs. Sequential Analysis for Population PK/PD Data I: Best–case Performance. *Journal of Pharmacokinetics and Pharmacodynamics*. **30(6)**:387–404.

[60] Zhang L, Beal SL, and Sheiner LB. (2003). Simultaneous vs. Sequential Analysis for Population PK/PD Data II: Robustness of Methods. *Journal of Pharmacokinetics and Pharmacodynamics*. **30(6)**: 405–416.

[61] Zhang L and Sheiner LB. (2005). Analyzing Multi–Response Data Using Forcing Functions. *Journal of Pharmacokinetics and Pharmacodynamics*, **32(2)**: 283–305.

Appendix A

A.1 BUGS code for Binary response model using Power Prior for Section 2.2

```
model{
for(i in 1:2)
{
Y[i]~dbin(p[i],n[i])
p[i]~dbeta((kappa*mu),(kappa*(1-mu)))
s[i]<-alpha0[i]
s2[i]<-(p[i]*alpha0[i])
}
sum1<-sum(s)
sum2<-sum(s2)
kappa<-2
mu<-0.5


thetahat<-sum2/sum1
kappastar<-kappa+sum1
mustar<-((kappa*mu)/kappastar) + (thetahat*sum1/kappastar)
ppb1<-(kappastar*mustar)
ppb2<- (kappastar*(1-mustar))
```

```
Y2~dbin(pp,np)
pp~dbeta(kappastar*mustar,(kappastar*(1-mustar)))
b1<-1
b2<-1
}
```

A.2 BUGS code for PK/PD Analysis for Section 3.2

```
## PK/PD WinBUGS code for the X-ALD (Lorenzos Oil) model
## NOTE: In what follows, DV2 = Y in the paper, and DV3 = Z in the paper
model{
for(i in 1:116){
endoerucic[i]  theta[2]
emax[i]  theta[3]
ec50[i]  theta[4]
c26endo[i]  theta[5]
eta1[i]dnorm(0,w[1])
eta2[i]dnorm(0,w[2])
eta3[i]dnorm(0,w[3])
eta4[i]dnorm(0,w[4])
eta5[i]dnorm(0,w[5])
eta6[i]dnorm(0,w[6])
for(k in offset[i]: (offset[i+1] ?1)) {
tvclerucic[k]  theta[1]*pow((WT[k]/70),0.75)
clerucic[k]  tvclerucic[k]
rateerucic[k]  DOSE[k] * (35.83)
cssex[k]
rateerucic[k]/clerucic[k]
cssav[k]  cssex[k]+endoerucic[i]
precDV2[k]  sigma[1]/cssav[k]*cssav[k]
DV2[k]dnorm(cssav[k], precDV2[k])
e[k]  emax[i]*cssex[k]
1038 C. BASU ET AL.
d[k]  ec50[i]+cssex[k]
```

```
effect[k]  e[k]/d[k]
diff[k]  c26endo[i]-effect[k]
precDV3[k]  sigma[2]/(diff[k]*diff[k])
DV3[k]dnorm(diff[k],precDV3[k])
}
}
theta[1]dnorm(42.3,t1)
theta[2]dnorm(0.582,t2)
theta[3]dnorm(0.655,t3)
theta[4]dnorm(3.34,t4)
theta[5]dnorm(0.971,t5)
for (j in 1:6) {w[j]  dgamma(1,1)}
for (j in 1:2) {sigma[j]  dgamma(0.1,0.1)}
} # end of BUGS code
## INITS:
list(theta = c(42, 0.5, 0.655, 3.34, 0.971),
w = c(1.228, 0.6849, 11.037, 20.61, 0.432, 33.44),
sigma=c(3,43))
## DATA: (NOTE: there are TWO loads, the list here,
and the primary data matrix below)
list(t1=0.04, t2=100, t3=100, t4=0.01, t5=100,
offset=c(
1, 17, 62, 103, 132, 153, 175, 184, 225, 259,
286, 306, 322, 357, 391, 431, 465, 508, 523, 535,
589, 619, 627, 631, 670, 707, 748, 766, 802, 847,
890, 925, 957, 965, 971, 1007, 1032, 1053, 1093,
1119, 1138, 1165, 1203, 1248, 1256, 1277, 1283,
1315, 1356, 1390, 1408, 1436, 1467, 1505,
1543, 1556, 1559, 1604, 1611, 1635, 1650, 1654,
1660, 1689, 1716, 1741, 1766, 1799, 1822, 1847,
1864, 1875, 1893, 1909, 1930, 1951, 1975, 1999,
2002, 2023, 2027, 2053, 2084, 2111, 2124, 2147,
2166, 2182, 2194, 2219, 2248, 2257, 2270, 2285,
2306, 2321, 2330, 2347, 2358, 2371, 2387, 2402,
2409, 2415, 2422, 2433, 2444, 2453, 2468, 2483,
```

```
2497, 2514, 2523, 2530, 2541, 2542, 2550))
## Here is the primary data matrix:
WT[] DOSE[] DV2[] DV3[]
19.5 0.0 0.59 0.46
19.5 9.0 85.19 0.41
19.5 9.0 10.85 0.35
19.5 9.0 5.45 0.45
19.5 9.0 18.47 0.29
19.5 9.0 14.90 0.24
19.5 9.0 16.50 0.40
19.5 9.0 35.75 0.28
22.8 9.4 50.72 0.36
22.8 9.4 11.76 0.30
22.8 9.4 9.17 0.43
22.8 9.4 71.82 0.51
25.4 9.8 2.05 0.58
. . .. . ..[snip]. . .. . ..
35.2 10.8 24.77 0.45
END
```

A.3 BUGS Code for Joint Clinical Efficacy and Toxicity Model for Section 4.4

Here is the BUGS code used to fit the joint power prior linear model for the Cincalcet data:

```
model
{

C <-100000

#individual random errors (assume only random slopes)
for (i in 1:Ntot)
{
mueta[i,1] <- eta0 + kdrug[i]*mud
```

```
mueta[i,2] <- eta0tox + kdrug[i]*mudtox
eta0_subj[i,1:2] ~ dmt(mueta[i,], tau_eta0[,],6)

}

const <- -0.5/(1-pow(rho,2))


##Pediatric
for (i in 1:N1)
{
for (j in 1:Tmax[i])
{

muY[i,j] <- eta0_subj[i,1]*t[i,j]
muYtox[i,j] <- thetaC + eta0_subj[i,2]*t[i,j]

logL[i,j] <-  0.5 * log(tau.e*tau.etox) - 0.5*log(1-pow(rho,2))
+const*(Y[i,j]-muY[i,j])*(Y[i,j]-muY[i,j])*tau.e
- const*2*rho*(Y[i,j]-muY[i,j])*(Ytox[i,j]-muYtox[i,j])
*pow((tau.e*tau.etox),0.5)
+ const*(Ytox[i,j]-muYtox[i,j])*(Ytox[i,j]-muYtox[i,j])*tau.etox

}

muYtox24[i] <- thetaC + eta0_subj[i,2]*24

muY24[i] <-  eta0_subj[i,1]*24
ptox24[i]<-step(8.4-muYtox24[i])
ptox2475[i]<-step(7.5-muYtox24[i])
peff24[i]<-1-step(30+muY24[i])
peff2420[i]<-1-step(20+muY24[i])
peff2450[i]<-1-step(50+muY24[i])

l[i] <- alphaC*sum(logL[i,1:Tmax[i]])
```

```
phi[i]<-  - l[i]+C
zero[i]~dpois(phi[i]);  useless[i] <- zerotox[i]



}

###Adults
for (i in (N1+1):Ntot)
{
for (j in 1:Tmax[i])
{
muY[i,j] <- eta0_subj[i,1]*t[i,j]
muYtox[i,j] <- thetaA + eta0_subj[i,2]*t[i,j]

logL[i,j] <-  0.5 * log(tau.e*tau.etox) - 0.5*log(1-pow(rho,2))
+const*(Y[i,j]-muY[i,j])*(Y[i,j]-muY[i,j])*tau.e
- const*2*rho*(Y[i,j]-muY[i,j])*(Ytox[i,j]-muYtox[i,j])
*pow((tau.e*tau.etox),0.5)
+ const*(Ytox[i,j]-muYtox[i,j])*(Ytox[i,j]-muYtox[i,j])*tau.etox

}

l[i] <- alphaA*sum(logL[i,1:Tmax[i]])
phi[i]<-  - l[i]+C
zero[i]~dpois(phi[i]);  useless[i] <- zerotox[i]
}

#Priors
eta0 ~ dnorm(0,0.01)
eta0tox ~ dnorm(0,0.01)

Omega[1,1] <- 1
  Omega[2,2] <- 1
  Omega[1,2] <- 0
Omega[2,1] <- 0
```

```
tau_eta0[1:2, 1:2]~ dwish(Omega[,], 2)

tau.e~dgamma(1,1)
tau.etox~ dgamma(1,1)
rho~dunif(-0.9,0.9)

thetaC~dunif(0,20)
thetaA~dunif(0,20)

mud ~ dnorm(0,0.01)
mudtox ~ dnorm(0,0.01)

drugslope <- mud + eta0
drugslopetox <- mudtox + eta0tox
}
```